The role of Nodule Cysteine Rich secreted peptides in controlling nodulation

by

Mingkee Achom

Thesis

Submitted to the University of Warwick
for the degree of

Doctor of Philosophy

School of Life Sciences

Supervisors: Miriam Gifford and Sascha Ott

September 2018

WARWICK
THE UNIVERSITY OF WARWICK

Dedication

This thesis is dedicated to my beloved parents

ema Heigrujam Shattabala Devi and *epa* Achom Khogen.

Also to my grandmother who sadly passed away on 18/12/18.

Contents

List of figures

List of Tables

Acknowledgments

First and foremost very special thanks to Miriam Gifford for agreeing to supervise me for my PhD. The support and encouragement I got from her during my PhD is immeasurable.

With a background of limited exposure to computational and systems biology, I joined her lab to work on this project. However, my sincere gratitude of thanks to her endless patience for believing in me. During my first year of PhD I was unaware of how I should start working with the data given to me and wanted to do lab work. However, that gradually changed towards the end as I realise the importance of computation in solving biological problems. And it is because of her, that I actually started enjoying such work that I used to avoid before.

I will also be forever thankful to Miriam for how she kept supporting me at thesis write-up stage. Her positivity and quick-response always motivate me not to leave things half-done. I think she has prepared me well not only to write and how to conduct quality research but also on how to face my next professional move.

In this context, the contribution of my co-supervisor Sascha Ott with his insightful comments and suggestions are valuable. I sincerely appreciate his expertise on the subject and find him inspirational. More than an hour long lab regular meeting with Sascha and Miriam over the last three years has profoundly changed to think about research approach. I will be always indebted towards their excellent supervision.

My fellow lab members of Gifford lab group has made this three year long journey of research seem short. Beatriz Lagunas great with her friendly guidance when I joined the lab as a new member. She was always there for me for experimental advice and my heartfelt thanks for all the help she gave.
I am also thankful to Alonso Pardal, Rana MF Hussain, Matt Teft, Liam Walker, Cantug Bar, Peter Morrison and Proyash Roy, Alexandra Grigogriadou for their several help. My sincere thanks to Gary Grant for his help and all the soil pots he prepared for my plants. Heartfelt thanks to Chitra who has helped me with all my last minute requests.

Last but not the least I thank a very dear person in my life, O. R. Metei for keeping me grounded, hopeful and cheerful.

Also to my two younger siblings Roger and Shneha, thank you for understanding me. Finally, special gratitude of thanks goes to my humble parents who never got to see what kind of research I actually did here. But nevertheless, their consistent support has made me reach until this stage. And I dedicate this work to them.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for a degree.

All research and analyses was carried out by the author except that the microarray hybridisation experiments in Chapter 3 were performed by Roxanna Bonyadi Pour, a previous PhD student in the Gifford lab; analysis of that data was performed by the author. The transcription factor search in Chapter 5 was performed in collaboration with Charlotte Rich, a PhD student in Patrick Schäfer lab's, University of Warwick.

A substantial part of Chapter 3 in this thesis has been submitted for publication in *Molecular Plant* and is presently in revision following review (Lagunas, Achom *et al*., 2019).

Abstract

Legumes house nitrogen-fixing endosymbiotic rhizobia in specialised polyploid cells within root nodules. The model legume *Medicago truncatula*, one of the members of the inverted repeat lacking clade (IRLC) of legumes is known to have a class of multi-gene family encoding for NCRs (Nodule Cysteine Rich peptides). Until now, studies have suggested that this large family acts as anti-microbial defensins keeping the rhizobial population in balance. However, these functional insights come from only a few of more than 500 NCRs that actually have diversified spatio-temporal expression. Due to their occurrence in large numbers from genome amplification, it is possible that these NCR genes have evolved gain of novel functions beyond just bacterial regulation.

Microarray transcriptomic data and promoter sequences revealed six novel conserved promoter motifs that are over-represented in subsets of NCRs that are regulated in different ways by combinations of rhizobia and nitrogen. Our hypothesis is that the NCR promoter motifs may be acting as transcription factor (TF) binding sites for the regulation of nodulation. Using the Arabidopsis TF database, we identified *CCA1, RVE1, ATHB15, ATHB16, AHL20* and *AHL25* as putative TF regulators. Possible existence of a nodule circadian clock in the *Medicago truncatula*-rhizobium symbiosis has been investigated and altogether our results support an expanded role of NCRs in signalling and nodule development during symbiosis.

Abbreviations and Symbols

-Rhiz	Mock inoculated
+Rhiz	Rhizobia inoculated
ABA	Abscisic acid
ACC	Ethylene precursor 1-aminocyclopropane carboxylic acid
AHL	AT-hook motif containing nuclear localised
AMLH	A17 Mock low to high N
AMT	Ammonium transport
ARLH	A17 Rhizobia low to high N
ARLL	A17 Rhizobia low to low N
AON	Autoregulation of nodulation
AVG	L-α-(2-AminoethoxyVinyl)Glycine hydrochloride
ATHB	Homeobox Leucine zipper *Arabidopsis thaliana*
BLAST	Basic local alignment search tool
bp	Base pair
bZIP	Basic leucine zipper
BLASTP	Protein BLAST
BLASTN	Nucleotide BLAST
C	Carbon
cDNA	Complementary deoxyribonucleic acid
CCA1	Circadian clock-associated protein 1a
CCaMK	Calcium/calmodulin-dependent protein kinase
CLE	CLAVATA Endosperm surrounding region
CRE	Cytokinin receptor
CRP	Cysteine rich peptides
Ct	Threshold cycle
dpi	Days post inoculation
DE	Differentially expressed
DEFL	Defensin like family genes
DEG	Differentially expressed genes
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
DNE	*DIE NEUTRALIS*
DNF	Defective in nitrogen fixation
DTT	Dithiothreitol
EBI	European Bioinformatics Institute
EE	Evening element
ELF	Early flowering
EMSA	Electro mobility shift assay
ENOD	Early nodulins

ERF	Ethylene responsive transcription factor
ERN	ERF required for nodulation
FC	Fold change
FDR	False discovery rate
FIMO	Finding individual motif occurrences
FNB	Fast neutron bombardment
FST	Flanking sequence tag
GA	Gibberellin
GDH	Glutamine dehydrogenase
GRP	Glycine rich peptide
GS	Glutamine synthetase
GSP	Gene specific primer
HAR	Hypernodulation aberrant root formation
HMM	Hidden Markov model
IRLC	Inverted repeat lacking clade
JA	Jasmonic acid
JCVI	John Craig Venter Institute
LBD	LOB DOMAIN-CONTAINING PROTEIN
LCO	Lipochitooligosaccharides
LHY	Late-elongated hypocotyl
LOB	Lateral organ boundary
LR	Lateral root
LRR-RLK	Leucine-rich repeat receptor-like kinase
LysM	Lysine motif receptor-like kinase
MADS	(for MINICHROMOSOME MAINTENANCE1, AGAMOUS, DEFICIENS and SERUM RESPONSE FACTOR)
MAFFT	Multiple alignment using fast Fourier transform
MFM	Modified Fahräeus medium
mRNA	Messenger ribonucleic acid
MSA	Multiple sequence alignments
MtDMI2	*Medicago truncatula* doesn't make infection 2
Myb	myeloblastosis family of transcription factors
N	Nitrogen
NAA	1-Naphthaleneacetic acid
NARK	Nodule autoregulation receptor kinase
NCR	Nodule cysteine rich secreted peptide
NF	Nod factor
NF	Noble Foundation, US
NFP	Nod factor receptor
NIN	Nodule inception protein
NiR	Nitrite reductase
NLP	Nodule inception protein-like protein

NR (NIA)	Nitrate reductase
NRT	Nitrate transporter family protein
NSF	Nitrogen fixation specificity
NSP	Nodulation signaling pathway
P	Phosphorus
PCR	Polymerase chain reaction
PEND	Plastid envelope DNA
PIP	Plasma membrane intrinsic protein
PR	Primary root
PRR	Pseudo response factor
PSSM	Position-specific scoring matrix
PWM	Position weight matrix
qPCR	Quantitative real-time polymerase chain reaction
r	Pearson correlation coefficient
RL	Rhizobium-legume symbiosis
RMA	Robust multiarray averaging
RNA	Ribonucleic acid
rpm	Revolutions per minute
RSA	Root system architecture
RVE1	Reveille-1 *Arabidopsis thaliana*
SLAC/SLAH	Slow anion channels homologs
SMLH	sunn-1 mock low to high N
SMLL	sunn-1 mock low to low N
SRLH	sunn-1 rhizobia low to high N
SRLL	sunn-1 rhizobia low to low N
SPL	Squamosa promoter binding protein-like
SUNN	Super numeric nodules
TF	Transcription factor
TFBS	Transcription factor binding site
Tnt	Transposable element *Nicotiana tabacum*
TOC1	Timing of CAB expression 1
UTR	Untranslated region
WOX	WUS homeobox-containing (WOX) protein family
WT	Wild type
WTS	Wild type sibling

Chapter 1: Introduction

1.1 Nitrogen and its uptake system in plants

Nitrogen (N) is an essential macronutrient that is required for formation of chlorophyll, amino acids, nucleic acids, and secondary metabolites. Plants acquire N available in the soil in the form of nitrate (NO_3) and ammonium (NH_4). N acquisition by the roots are transported in the plant system or assimilated with carbon to form amino acids (Miller et al., 2007). Both N uptake and metabolism is required for proper metabolism, growth and development of the plants (Bouguyon et al., 2012; Ruffel et al., 2008; Vidal and Gutierrez, 2008).

However, soil N availability is not evenly distributed due to factors such as soil leaching or uptake by plant roots that affect the root system environment and its associated microbiome (rhizosphere) in the soil (Hirel et al., 2011). Such fluctuations in the soil nitrogen content are thought to have driven plant evolution of complex nutrient uptake, transport and assimilation systems for adaptation to the external environment (Canales et al., 2014; Gutierrez, 2012).

NO_3^- acquired by root cells enters the assimilatory pathway for the reduction of nitrogen first into nitrite (NO_2^-) and then to ammonium (NH_4^-), catalysed by the two enzymes nitrite and nitrate reductase respectively (Bouguyon et al). Two major enzymes, glutamine synthetase and glutamate synthase are involved in the subsequent assimilation of ammonium into amino acids (Bouguyon et al., 2012; Lam et al., 1996). Unlike in non-legumes, nitrogen present in the atmosphere is converted to ammonia by direct fixation in the root nodules of legume plants (Lam et al., 1996).

To enable the optimal acquisition, utilization and transport of NO_3^-, plants need to detect the external environment. Such sensory and signalling action of NO_3^- for detection can be seen in the induction of genes that encode for nitrite and nitrate reductase, nitrate transporter family proteins NRT1 or NRT2 (Lejay et al., 1999; Wang et al., 2004).

Nitrogen responses, transport and its metabolism in plants have been well studied, mostly in the model plant *Arabidopsis thaliana* (O'Brien et al., 2016). At least four nitrate transporter families - NPF, NRT2, CLC, and SLAC/SLAH have been identified in plants (Krapp et al., 2014). NO_3 uptake and translocation in plants involves complex regulation and cross-talk between four main types of transporters, NPF (NRT1/PTR), NRT2, CLC, and SLAC1/SLAH (Wang et al., 2012). Among these transporters, two main classes of transporters, NPF (formerly known as 'low-affinity' NRT1/PTR) (Leran et al., 2014) and high-affinity NRT2 in Arabidopsis (Tsay et al., 2007) have been well characterized (Figure 1.1) (O'Brien et al., 2016). Low affinity transport system (LATS) functions at high level of N concentrations (>1 mM) while high affinity transport system (HATS) functions at low N (μM) concentrations (Kraiser et al., 2011); thus the N transport systems can respond effectively at many N levels.

Although the NPF members mainly function as low affinity at high N concentrations, some NPF proteins, such as NRT1.1 (also named NPF6.3 or CHL1) in Arabidopsis (Tsay et al., 2007) and MtNRT1.3 in the model legume *Medicago truncatula* (Morere-Le Paven et al., 2011) act as dual-affinity transporters involved in both low and high affinity transport system.

Apart from transporter family proteins, transcription factors (TFs) of the nitrate signalling pathway have been identified as key regulators for NO_3^- responsive genes (Castaings et al., 2009; Vidal et al., 2015). One such TF regulator is the NIN like protein 7 (NLP7) TF that shares homology with NIN (Nodule Inception) protein, one of the key regulators in legume nodulation (Schauser et al., 2005). NLP7 is involved in the direct interaction target gene promoters as a regulator of the primary nitrate response and signalling pathways (Castaings et al., 2009; Marchive et al., 2013). Other key TFs involved in the NO_3^- response regulator include TGACG MOTIF-BINDING FACTOR (TGA1),TGA4, ARABIDOPSIS NITRATE REGULATED 1 (ANR1), BASIC LEUCINE-ZIPPER 1 (bZIP1), LOB DOMAIN-CONTAINING PROTEIN (LBD37), LBD38 and SQUAMOSA PROMOTER BINDING PROTEINLIKE 9 (SPL9) (O'Brien et al., 2016).

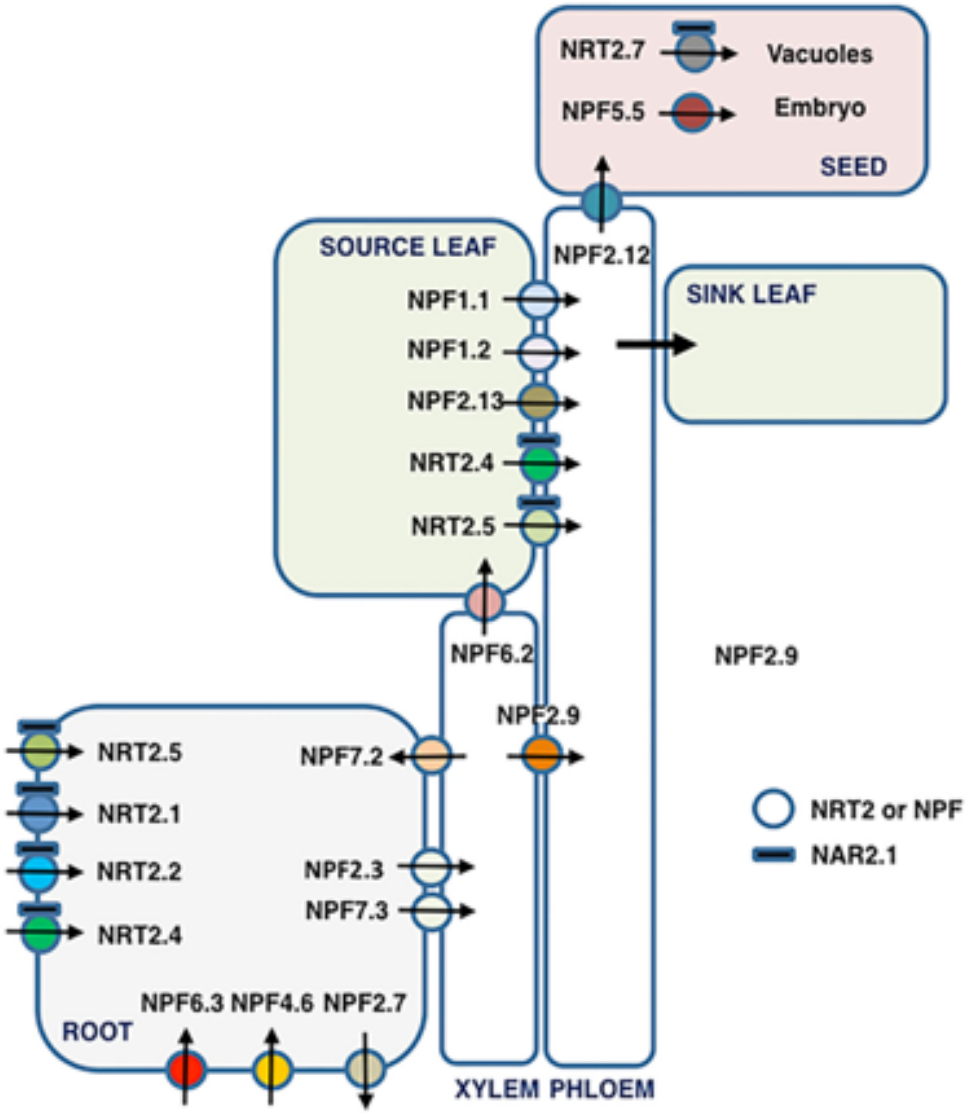


Figure 1.1 NRT2 and NPF families of *Arabidopsis* nitrate transporters. The functions depicted are: root uptake (influx/ efflux), loading/unloading of the xylem, loading/unloading of the phloem, accumulation in seed vacuoles, and transport into the embryo. At the cellular level, all proteins are localized at the plasma membrane, except NRT2.7, which is localized at the tonoplast (O'Brien et al., 2016)

In addition to NO_3^- , NH_4^+ is also taken up by plants. Supplied in mixed N forms of NO_3^- and NH_4^+, NH_4^+ is often the preferential form of N uptake by plants (Gazzarrini et al., 1999). In Arabidopsis, ammonium transport (AMT) proteins such as AtAMT1;1, AtAMT1;2, and AtAMT1;3 play a major role in N uptake in the form of NH_4^+ (Gazzarrini et al., 1999). Ammonium transport by the protein family of ammonium transporter 1 (AMT1) including AMT1.1, 1.2, 1.3

and 1.5 (Alvarez et al., 2012) act as uniporters or NH_3/H^+ co-transporters (Kraiser et al., 2011) to import ammonium from the rhizosphere or atmosphere. This imported ammonium is assimilated by the action of two main enzymes, glutamine synthetase (GS) and glutamate dehydrogenase (GDH) (Hirel and Lea, 2001).

Ammonium is another preferred form of N uptake by N-fixing symbiotic bacteria in host legume root nodules (Udvardi and Day, 1997). NH_4^+ concentrations in the cytosol of legumes can be about 50-fold lower than endosymbiont bacteroids because ammonium transport in root nodules takes place across the symbiosome membrane (Streeter, 1989). Efficient uptake of ammonium in such cases requires low-affinity and high-capacity transport systems on the plant side of the symbiotic interaction (Tyerman et al., 1995).

Nutrient acquisition are strongly influenced by long-distance signalling in plants. Use of the split root system in *Arabidopsis* has demonstrated that a long-distance signalling mechanism is involved in the uptake and adaptive response of root system architecture (RSA) to a heterogenous supply of N (Mounier et al., 2014; Ruffel et al., 2011). For example, split roots displayed a differential lateral root development in high/low N environments (Gansel et al., 2001). Similar split-root system experiments with transcriptome analyses in *Medicago truncatula* have shown that NO_3^-, NH_4^+ and N_2- fixation acquisition systems do not have similar responses to changes in plant N status (Ruffel et al., 2008).

RSA displays a high level of plasticity in response to heterogeneous nutrient supply, including the source of N (Patterson et al., 2010) and its status (Wang et al., 2007). RSA can be defined as the size and growth direction(s) of the primary root (PR), lateral roots (LR) and root hairs. A high level of cell type (Gifford et al., 2008) and tissue type specificity (Wang et al., 2003a) underlies RSA responses. This is one aspect of plant root adaptation to rapidly changing environments (Alvarez et al., 2012; Canales et al., 2014; Gutierrez, 2012).

1.2 N-fixation by symbiotic bacteria

Legumes benefit from nodule formation with the symbiotic association between the soil rhizobia and host root enabling uptake of biologically fixed nitrogen. Thus, legumes are one of the important crop plants that helps in soil

nitrogen enrichment and they are often used in crop rotation (Iannetta et al., 2016; Stagnari et al., 2017).

In legumes, available atmospheric N can be taken up by plants once it is fixed by soil borne bacteria collectively called as rhizobia. These endosymbiotic nitrogen-fixing rhizobia are housed in specialised polyploid cells called nodules (Maroti and Kondorosi, 2014; Sprent and James, 2007). The nodule symbiotic plant cells are polyploid because of several cycles of endoreduplication (genome replication with cessation of later stage of mitotic division) and enlarged to accommodate the rhizobia (Mergaert et al., 2006; Sprent and James, 2007).

Nodulation process starts with signal exchange between the rhizobia (that releases Nod Factors, NF as signals) and host plant roots (secrete flavonoids in turn) (Figure 1.2). NF reception involving lysine-motif domain (LysM) activates calcium oscillations which regulate gene expression in both the host legume and rhizobia to enable the establishment of symbiotic relationship. Thereby, colonisation of the plant root is enabled by the formation of an infection thread that allows controlled bacterial entry (Figure 1.2).

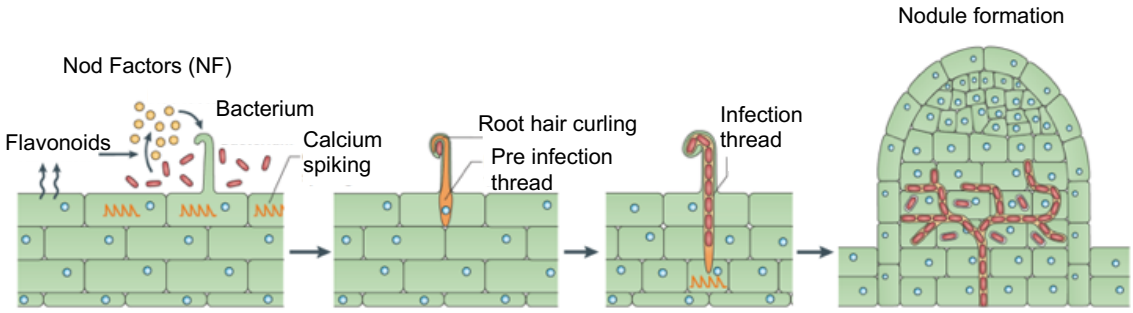


Figure 1.2: Early events during nodulation. Plant roots released flavonoids, thereby signalling rhizobia in the rhizosphere to produce nodulation factors (Nod factors, NF) that are in turn recognised by the plant. NF perception activates the symbiosis signalling pathway, leading to calcium oscillations, cell division in the cortex to form a nodule primordia. Rhizobia enters the plant root via root hair curling and from this site, infection threads are formed, finally progressing to the inner cortical cells and the nodule primordium (Oldroyd, 2013).

1.3 Nodulation and lateral root development

Nodulation as such consists of two closely coordinated processes: (i) the organogenic process, in which the nodule tissue is formed to accommodate the bacteria and (ii) the infection process, in which bacteria colonises inside the host plant (Madsen et al., 2010). Genetic and molecular analysis suggests that nodule development integrates pre-existing plant regulatory pathways that are related to root organogenesis (Mathesius, 2003). For example, nodulation and lateral root (LR) development share a common environmental regulation controlled by hormone signalling that in turn influences N fixation (De Smet et al., 2007; Lopez-Bucio et al., 2003; Tirichine et al., 2007).

Nodules and LR are both induced during low N in the presence of rhizobium, and the morphological development of these two organs is fundamentally linked with accumulating evidence that symbiotic organisms can affect root architecture (Maillet et al., 2011; Olah et al., 2005; Raven and Edwards, 2001). LRs have known to be in existence since at least 400 million years ago (Raven and Edwards, 2001). Nodules, however, have evolved more recently, around 60 million years ago, driven by N deplete and CO_2 rich environmental conditions (Sprent, 2007). Thus, a current hypothesis is that mechanisms that regulate nodule development may have been co-opted from the existing processes that regulate root development pathway (Hirsch et al., 1997).

In conjunction with the tight regulation of developmental pathways, one notable overlap is the shared mechanism that involves hormonal control of nodulation and root system architecture (Hirsch et al., 1997; Oldroyd, 2013; Yendrek et al., 2010). Several regulators of cross-talk between nodulation and LR development have been characterised. For example, NFs as rhizobial signalling molecules have a crucial role in the symbiosis as well as LR induction. LRs have shown to be stimulated by NFs and many of the same genes (e.g. *NFP*, *DMI1*, *DMI2*, *DMI3* and *NSP1*) required for symbiotic response (Olah et al., 2005).

1.4 Phytohormonal roles during nodulation and LR development

LR and nodules are two lateral organs that enable root plasticity in plants to allow adaption to external environment changes. However, both LR and nodule development are under the control of phytohormonal gradients, in particular auxin and cytokinin. Cytokinin signalling in legume roots also affects auxin signalling, as can be seen (Figure 1.4, (Oldroyd et al., 2011)) from CRE1-dependent changes in polar auxin transport (Grunewald et al., 2009; van Noorden et al., 2006). Such locally induced changes in polar auxin transport is crucial because it inhibits accumulation of auxin at the site of bacterial infection and initiates nodule organogenesis (Hirsch et al., 1989; Mathesius et al., 1998).

The interplay of both auxin and cytokinin is important not only for nodule initiation (Frugier et al., 2008) but also for lateral root primordia development (Benkova and Bielach, 2010; Peret et al., 2009). Auxin localisation in root pericycle cells drives LR meristem initiation. But after initiation, auxin levels are maintained through the activity of positive regulator AUX1 (Figure 1.3, (Oldroyd et al., 2011) and the inhibition of negative auxin regulators (Peret et al., 2009). At the LR primordial stage, cytokinin acts as LR initiation inhibitor, but once initiated LR insensitivity to cytokinin is marked by strong cytokinin induction in the mature LR itself (Benkova and Bielach, 2010; Peret et al., 2009).

While cytokinin appears to inhibit LR primordia initiation, the opposite mechanism is seen at nodule initiation (Frugier et al., 2008). At the nodule meristem initiation stage, cytokinin acts as primary activator and regulator of polar auxin transport for nodule organogenesis (Frugier et al., 2008). At this level of regulation then, nodules and lateral roots share the same genetic machinery but modulation of auxin and cytokinin levels have opposite outcomes. Hence, high localisation of cytokinin and low auxin levels favour nodule organogenesis (Oldroyd et al., 2011).

Aside from primary control of LR organogenesis by auxin and cytokinin, ethylene and abscisic acid (ABA) have been noted as negative regulators of nodule numbers by affecting calcium signalling (Mortier et al., 2012b). Ethylene is reported to inhibit root hair curling (Gresshoff et al., 2009; Lohar et al., 2009), reduce calcium spiking and repress early nodulins, ENODs (Oldroyd et al.,

2001). Nodulation at later stages is then disrupted by ABA that represses cytokinin signalling, affects NF signalling pathways and calcium oscillation (Ding et al., 2008)

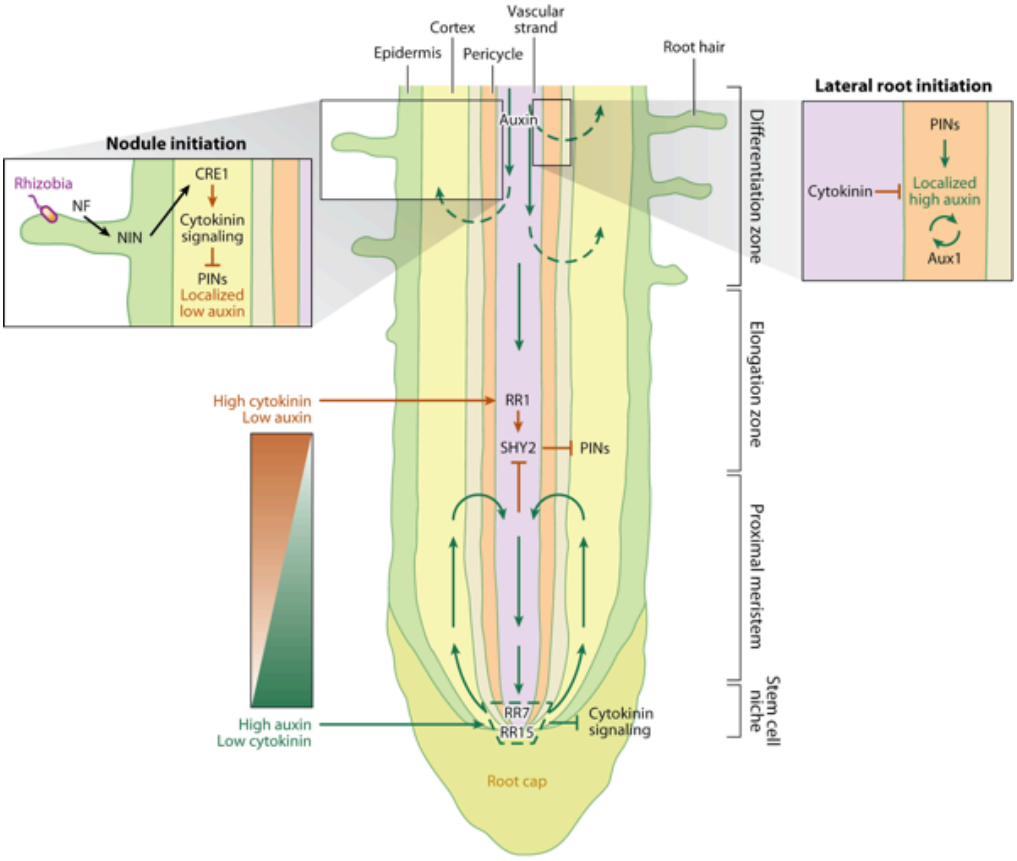


Figure 1.4: The roles of auxin and cytokinin during root meristem development. Apical meristem division into 4 zones can be seen; stem cell niche where the quiescent center resides, the proximal meristem where cell divisions occur, the elongation zone and the differentiation zone where the specialized morphology and functions of cells develop. Auxin transport proteins (PINs) traverse down the root and accumulate at the stem cell niche tip (green arrows). Auxin cycle flow reduction in the elongation zone, likely through cytokinin suppression of PINs (orange bar). Thereby an auxin-cytokinin gradient is established with high auxin at the stem cell niche and high cytokinin at the elongation zone. This antagonistic auxin-cytokinin gradient is maintained by auxin induction of RR7 and RR15 (that suppress cytokinin signalling) and cytokinin induction of SHY2 (that inhibits PIN expression). Localised accumulations of auxin in the pericycle marks the emergence of LR sites. Cytokinin suppress LR emergence by the suppression of auxin PINs. During nodule development, localized cytokinin signalling in the root cortex with a low level of auxin initiates the nodule primordia. Activation of nodule organogenesis requires Nod factor (NF) recognition at the root surface, and this induces the NIN transcription factor to induce cytokinin signalling in the cortex through the upregulation of *CRE1* (Oldroyd et al., 2011).

1.5 Autoregulation of nodulation (AON) in legumes

In addition to the localised phytohormonal control of nodule numbers, long distance signalling - autoregulation of nodulation (AON) - is an important determinant of nodule numbers. Split root experiments have shown AON to be a root derived nodulation signal that moves to the shoot and back again via a shoot derived signal that regulates nodulation and nodule numbers Figure 1.5 (Kassaw et al., 2015; Mortier et al., 2012b; Reid et al., 2011).

Autoregulation of nodulation can be divided into root- and shoot-derived aspects (Fig. 1.5). The root derived factors involve responses to initial cell divisions that lead to the induction of Q (cue signal) for nodulation. A second class of root-dependent components then act downstream to perceive the shoot signal that leads to inhibition of further nodule development (Reid et al., 2011).

AON activation occurs in the cortex after cytokinin signalling (Mortier et al., 2012b) and systemic inhibition of nodulation from AON thereby affects nodule primordial development (Li et al., 2009a; Suzuki et al., 2008). AON uses a feedback suppression mechanism from a root-derived signal that is thought to move via the phloem (Oka-Kira and Kawaguchi, 2006). Grafting experiments have shown that nitrate-induced CLE (NIC1) peptide is perceived by a root-localised CLAVATA1-like Leucine-Rich Repeat Receptor-Like Kinase (LRR-RLK) called Nodulation Autoregulation Receptor Kinase (*Gm*NARK) (Searle et al., 2003).

In *Lotus japonicus, Lj*CLE-RS1 and *Lj*CLE -RS2; (Okamoto et al., 2009) CLAVATA3/endosperm-surrounding region-related (CLE) peptides are involved in AON. In *Medicago truncatula* it involves *Mt*CLE12 and *Mt*CLE13 (Mortier et al., 2012a) and in soybean, *Gm*NIC1(Reid et al., 2011). *Mt*CLE12, *Mt*CLE13, *Lj*CLE-RS1 and *Lj*CLE-RS2 have been show to activated downstream of NIN shortly after nodule primordium initiation (Mortier et al., 2012a; Okamoto et al., 2009). Overexpression of these specific CLE peptides lead to nodule suppression in wild-type plants (Mortier et al., 2012a; Okamoto et al., 2009). Similar CLE peptides in *Arabidopsis* have been shown to be

involved in root elongation, vascular differentiation and developmental regulation of the shoot apical meristem (Strabala et al., 2006).

Several other mutants have been isolated that exhibit an increased nodulation phenotype, known as hyper- or supernodulation. Those functioning in the root include *rdh1* (Ishikawa et al., 2008), *rdn1* (Schnabel et al., 2011), *too much love* (Magori et al., 2009), *plenty* (Yoshida et al., 2010), *efd-1* (Vernie et al., 2008), *astray* (Nishimura et al., 2002), and *sickle* (Penmetsa et al., 2008). These mutants show aberrant nodulation with increased number of nodules and disruption in the AON pathway (Reid et al., 2011).

In *Lotus japonicus* this hypernodulant gene is encoded by *LjHAR1* (Hypernodulation Aberrant Root Formation 1) (Wopereis et al., 2000) and in *M. truncatula* by SUperNumary Nodules, *MtSUNN* (Schnabel et al., 2005). These mutants display disruption in the autoregulation of nodulation (AON) pathway, which consists of at least two systemic regulatory circuits to control nodule numbers and activity (Kassaw et al., 2015). *HAR-1* of *Lotus japonicus* and *SUNN* of *M. truncatula* are orthologous to *Arabidopsis CLAVATA1*(Krusell et al., 2002; Schnabel et al., 2005).

In *M. truncatula*, nitrate limitation has been demonstrated to result in both local and systemic regulation of nodulation that partially depend on *SUNN* (Jeudy et al. 2010). *mtsunn* mutants can nodulate in the presence of nitrate, indicating that a combinatorial action of autoregulation and a nitrate signal both inhibit nodule progression (Schnabel et al., 2005). Along with SUNN, RDN1 (Root-Determined Nodulation) controls AON responses in *M. truncatula*, and CLE peptides are thought to be involved as both the local and systemic nodulation status signal (Mortier et al., 2012a). In *L. japonicus* it was shown that *NITRATE UNRESPONSIVE SYMBIOSIS 1* (*NRSYM1*), a NIN-like gene, regulates nodule numbers and nodule development in response to nitrate levels (Nishida et al., 2018).

A similar observation has been noted with other nitrate tolerant phenotypes in the AON mutants *har1, nark, sym29, klv* and *tml*; they are able to nodulate in high nitrate conditions (Magori et al., 2009; Penmetsa et al., 2003; Schnabel et al., 2005; Wopereis et al., 2000). Such evidence implies a complex

cross talk between nitrate tolerance, AON and RSA development during the nodulation process.

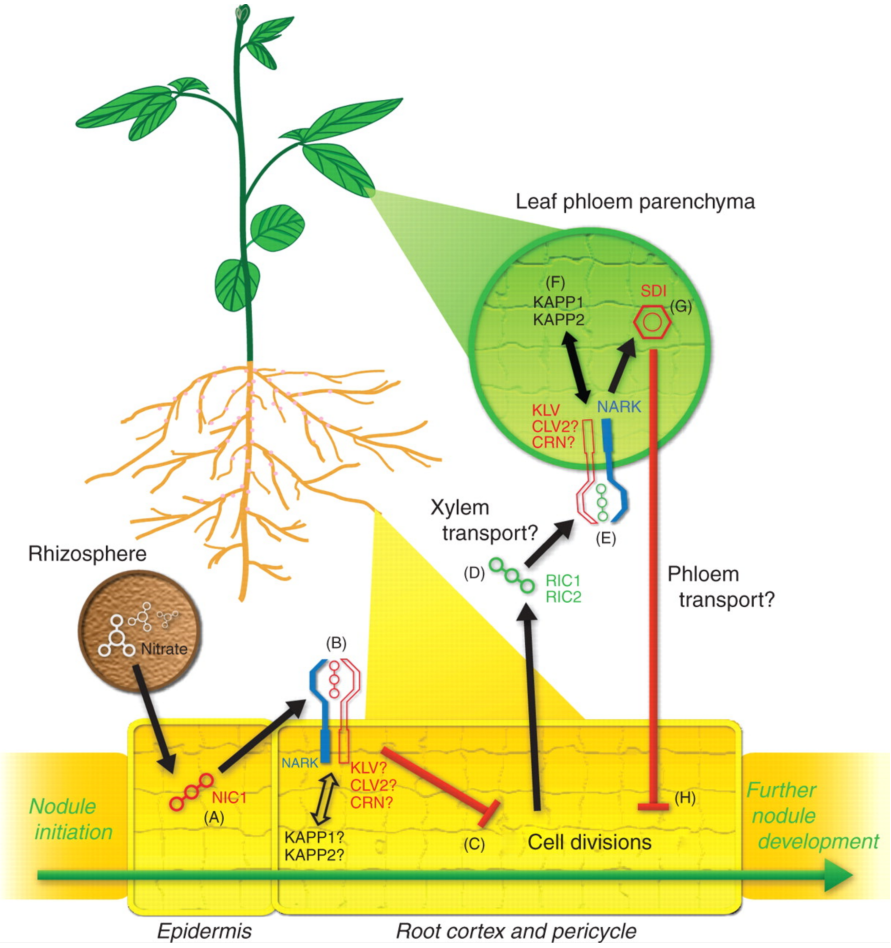


Figure 1.5 A working model of autoregulation of nodulation (AON) in legumes. Legumes regulate nodulation in response to pre-existing infections and soil nitrogen levels. Nitrate induces the production of CLE peptide (NIC1; A) that acts locally in the root via the AON receptor kinase, NARK (B; or its orthologues in other species), to suppress nodule development (C). NARK may act along with other components to perceive NIC1. Rhizobia-induced CLE peptides are induced at several stages of nodule development and may be to the shoot through xylem. (D) In the shoot, NARK and other possible receptor-like kinase proteins in *Arabidopsis* (CLV2, KLV and CRN (Jeong et al., 1999; Miyazawa et al., 2010; Muller et al., 2008) perceive these putative ligands (E). Two kinase-associated protein phosphatases (KAPP1/2) are phosphorylated by NARK and in turn dephosphorylate the NARK kinase (F). An equilibrium of phosphorylation between these components may be required preceding the production of the shoot-derived inhibitor (SDI; G). SDI is then transported via the phloem to the roots where it suppresses nodule progression and development. (H). A compound similar to SDI may be involved in the nitrate pathway that acts locally to inhibit the progression of nodule formation (Reid et al., 2011)

1.6 Endosymbiotic bacteroid development and nodule types in legumes

Rhizobial entry to the root begins with the chemical communication between the bacteria and the host plant roots. Recognition of NFs, rhizobia derived lipochito-oligosaccharides, by lysine motif receptor like kinases (LysMs) present in the plasma membrane of the root epidermal cells is a highly specific interaction due to structural variety of NF from different rhizobia (Bensmihen et al., 2011).

Structurally, NFs consist of a chitin backbone with two to six β-1,4-linked N-acetyl-D-glucosamine residues with N-linked fatty acid attached to the terminal sugar (Cooper, 2007). Nod factor recognition by the LysM receptors then initiates the signalling pathway activating transcriptional responses that control nodule organogenesis, rhizobial colonisation. This is essentially the first signal that leads to controlled rhizobial entry via an infection thread and development of host plant derived membranes called 'symbiosomes' (Kouchi et al., 2010).The symbiosome encapsulates the bacteria, separating them from the plant cytosol (Brewin, 2004; Perret et al., 2000). Bacteria inside the symbiosome then differentiate into nitrogen-fixing bacteroids that enable N supply to host plants in exchange for carbon and other nutrients (Gibson et al., 2008; Jones et al., 2007). Bacteroides under such circumstances undergo various morphological and metabolic changes to fix the atmospheric dinitrogen into ammonia (Udvardi and Poole, 2013). For example, in *Medicago truncatula*, endoreduplication of the *Sinorhizobium melliloti* takes place, increasing the chromosome pair count to 24 as compared to one-two in the free-living bacteria (Mergaert et al., 2006).

Despite the infection by same rhizobial strains, different legume species develop different nodule types (Mergaert et al., 2006). There are two different types of nodulation - indeterminate (nodules do not lose meristematic activity, e.g. *Pisum sativum*, *Medicago truncatula*, *Galega orientalis*) and determinate (round shaped and short-lived meristematic activity, e.g. *Glycine max, Lotus japonicus, Phaseolus*) (Sprent and James, 2007). Endoreduplication is a feature for a few host plant roots that form indeterminate nodules (Vinardell et

al., 2003). Indeterminate nodule type legumes belong to Inverted Repeat-Lacking Clade (IRLC) species, identified by the loss of a 25-kilobase inverted repeat in the chloroplast genome (Lavin et al., 1990; Pan and Wang, 2017).

Indeterminate nodules can be divided into different zones: (1) actively dividing meristematic zone; (2) infection zone, infection site of rhizobia and release into the cytoplasm; (3) interzone, zone of elongation and enlargement for the bacteroid differentiation (4) fixation zone, nitrogen fixing zone by matured bacteroids (Figure 1.6 C, (Pan and Wang, 2017).

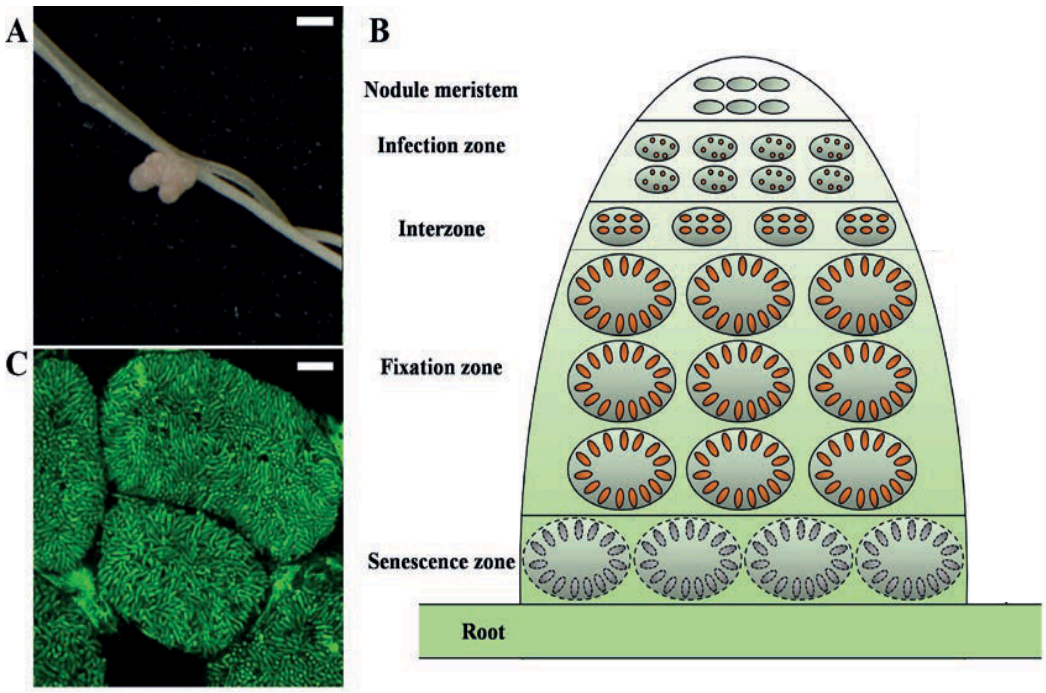


Figure 1.6 Differentiation of rhizobium bacteria inside Medicago nodules. (A) A representative picture of three-week-old indeterminate nodule type *M. truncatula* A17 nodules. Scale bar, 2 mm. (B) A representation of different type of meristematic zones in indeterminate nodules. (C) A confocal image of symbiotic nodule cells containing actively differentiating bacteroids. Rhizobia used for inoculation express GFP constitutively. Scale bar, 20 μm. (Pan and Wang, 2017)

1.7 Nodule cysteine rich peptides

The Inverted Repeat-Lacking Clade (IRLC) legume family including the model legume *M. truncatula* possess a strikingly large gene family encoding for Nodule Cysteine-Rich (NCR) peptides (Alunni et al., 2007; Mergaert et al., 2003). IRLC legumes mediate symbiosome at the cellular level via hundreds of

these Nodule Cysteine-Rich (NCR) peptides that control the activity of the rhizobial bacteria (Figure 1.7 A) to undergo terminal differentiation (Mergaert et al., 2006). NCRs have not been reported in determinate nodule forming species such as *Glycine max*, *Lotus japonicus* and *Phaseolus vulgaris* (Fedorova et al., 2002; Graham et al., 2004; Mergaert et al., 2003). There are findings that symbiotic cells of some *Aeschynomene* spp. belonging to more ancient lineage of the Dalbergioid clade also contain a large NCR-like peptide family with similar mechanisms to that of IRLC legumes (Czernic et al., 2015).

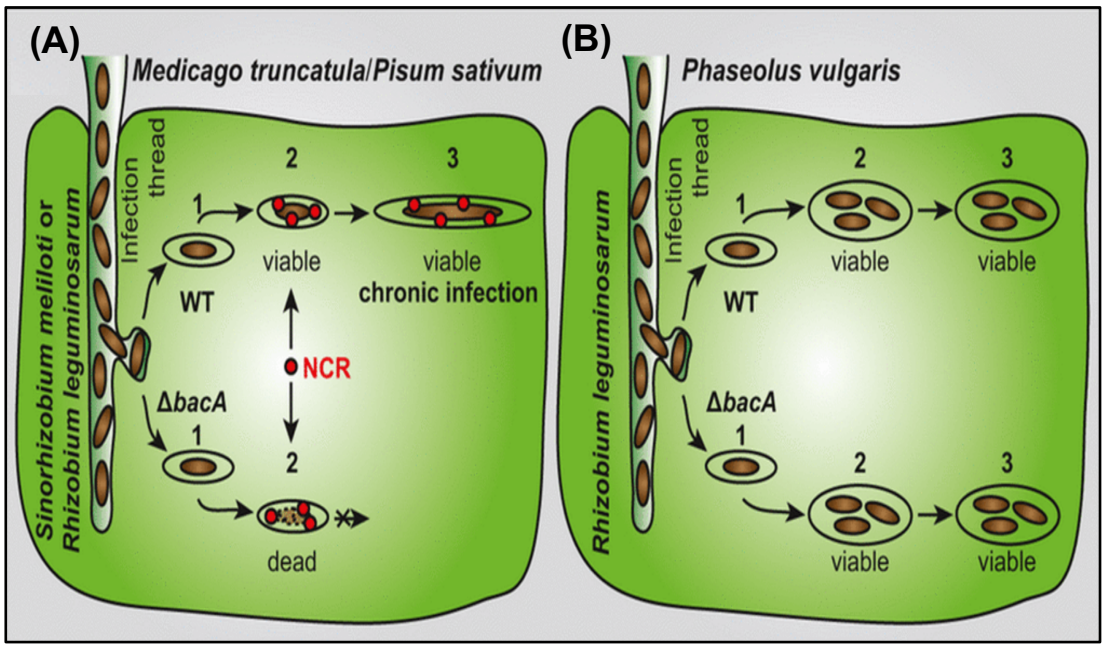


Figure 1.7: Rhizobial interaction in legume root and nodule cells. (A) legumes of the IRLC clade such as *M. truncatula* and *P. sativum*, rhizobia terminal differentiation is regulated by NCR peptides (3). The bacterial BacA protein is essential for protecting the rhizobia against the antimicrobial activity of NCR peptides. (B) In contrast, non-IRLC legumes such as *Phaseolus vulgaris* do not produce NCR peptides and BacA is dispensable for rhizobia. Rhizobia do not exhibit terminal differentiation in such host legumes and often multiple bacteroids can be found inside a single symbiosome membrane (Haag et al., 2013).

The *M. truncatula* family of NCRs consists of more than 500 genes and is reported to have different spatio-temporal expression patterns but highly specific to the nodule region only (Guefrachi et al., 2014; Nallu et al., 2013). However, only eight NCRs have been functionally characterised at the transcriptomic level and confirmed to act as antimicrobial peptides that

resemble defensins (Tesfaye et al., 2013). Proteomic studies revealed that NCRs contain a signal peptide and the secreted mature peptides have conserved cysteine patterns; this feature does differentiate them from defensins (Durgo et al., 2015).

Several studies using genetic approaches have demonstrated that host protein secretion has been to induce terminal differentiation of nodule bacteroids. A study of the nodule-specific signal peptidase complex of *Medicago DNF1* (*Defective In Nitrogen Fixation 1*), showed that the maturation of secretory proteins and peptides is critical for the bacteroid differentiation (Wang et al., 2010). Obstruction of NCR transport in the *dnf1-1* signal peptidase mutant correlated with the absence of terminal bacterial differentiation (Van de Velde et al., 2010). In another study (Horvath et al., 2015), it was reported that inactivation of a single *NCR* gene (NCR169) led to the early senescence of symbiotic nitrogen fixation in Medicago. However, these recent reports on studies in the functional analysis of NCRs during nodulation is still in its infancy representing only a handful of the functional activity of this large gene family.

Identifying *in vitro* antimicrobial activities of cationic NCRs (Ordogh et al., 2014; Tiricz et al., 2013) have indicated that NCRs resemble defensin type antimicrobial peptides. Such NCRs partly disturb the microbial membranes (Mikulass et al., 2016) or interact with proteins involved in transcription, translation, and cell division in the *Sinorhizobium meliloti-M. truncatula* symbiosis (Farkas et al., 2014; Penterman et al., 2014; Van de Velde et al., 2010).

Alongside the role of NCRs in defense responses in rhizobium-legume symbiosis, there are findings that cysteine rich peptides can also function as signalling peptides regulating the root developmental pathways (Marshall et al., 2011). The presence of the highly conserved 4-6 cysteine residues in the NCRs therefore provides a strong case to investigate if NCRs have additional roles beyond anti-microbial functions.

Evolutionary lineages in legumes have been derived from a common ancestor 60 million years ago (Lavin et al., 2005). In the Papilionoid clade including *M. truncatula*, evolution of nodulation for endosymbiotic nitrogen fixation has been greatly shaped by the whole genome duplication event

approximately 58 mya (Cannon et al., 2010). *M. truncatula* gene duplication is 3.1-fold higher than in *G. max* and1.6-fold higher than in *Arabidopsis* (Young et al., 2011). The excess local gene duplications in *M. truncatula* occurs genome-wide and affects many gene families (Young et al., 2011). This appears to have resulted in amplification and rearrangement of many gene families (e.g. F-box gene families, nucleotide-binding site and leucine-rich repeat NBS-LRR) and retention of paralogous genes (*NOD FACTOR RECEPTOR, NFP* and *ETHYLENE RESPONSIVE FACTOR FOR NODULATION*, *ERN1*) (Young et al., 2011).

Among the amplified gene families, the *NCR* gene family of *M. truncatula* and its close relatives with indeterminate nodule-type (Kato et al., 2002) is noteworthy (Young et al., 2011). The rapid expansion of the NCRs is thought to be associated with gain of nodulation function in legumes (Nallu et al., 2013). With the emergence of phylogenetic evidence that local gene duplications shape gene family expansion (e.g. F-box family; Young et al., 2011) in Medicago, and the discovery that some nodule related genes have co-evolved from their ancestor and conserved (Young et al., 2011) it can be hypothesised that NCRs play more diverse roles than just as defensins. Consistent with the reshaping of the genome during the course of evolutionary events and the broad-scale difference observed in spatio-temporal expression profile of NCRs, the question is raised, if *NCR* genes have co-evolved from pre-existing ancient gene families. Hence, we know relatively little about if the roles of *NCR* genes expanded after the numbers of *NCR* genes expanded.

1.8 The circadian system in *Arabidopsis* and legume plants

Circadian systems are widespread endogenous mechanisms that allow organisms to co-ordinate daily temporal cycles of day and night in accordance to their physiological changes. Their wide distributions have evolved in several organisms, from prokaryotic cyanobacteria to mammals, indicating their importance in biological processes. Among the extensive pathways of 24 h oscillatory rhythms that are under the circadian system; nitrogen-fixation in cyanobacteria, olfactory responses in Drosophila and sleep patterns in humans can be mentioned (Young and Kay, 2001).

In plants, interaction with the environment heavily relies on the periodic circadian rhythm for effective output and performance. These internal clock-generated biological rhythms allow the plants to anticipate environmental changes, resulting in regulation of physiological and developmental adaptation. The endogenous clock in plants and circadian rhythms are involved in control of numerous physiological processes (Figure 1.8; Yakir et al., 2007) including photosynthesis, leaf movement, hormone responses, stem elongation and stomatal opening but also root development (Harmer, 2009; Michael and McClung, 2003; Pruneda-Paz and Kay, 2010).

Gene expression in plants is also tuned to periodic circadian rhythms to optimize performance relative to light/dark periods when grown in day/night cycles. Much of our understanding of plant circadian clock comes from the model plant *Arabidopsis* (Yakir et al., 2007). Transcriptomic investigation of the *Arabidopsis* genome has indicated that about one-third of transcripts are under circadian regulation (Covington et al., 2008; Michael and McClung, 2003), and the case is likely to be similar in other species.

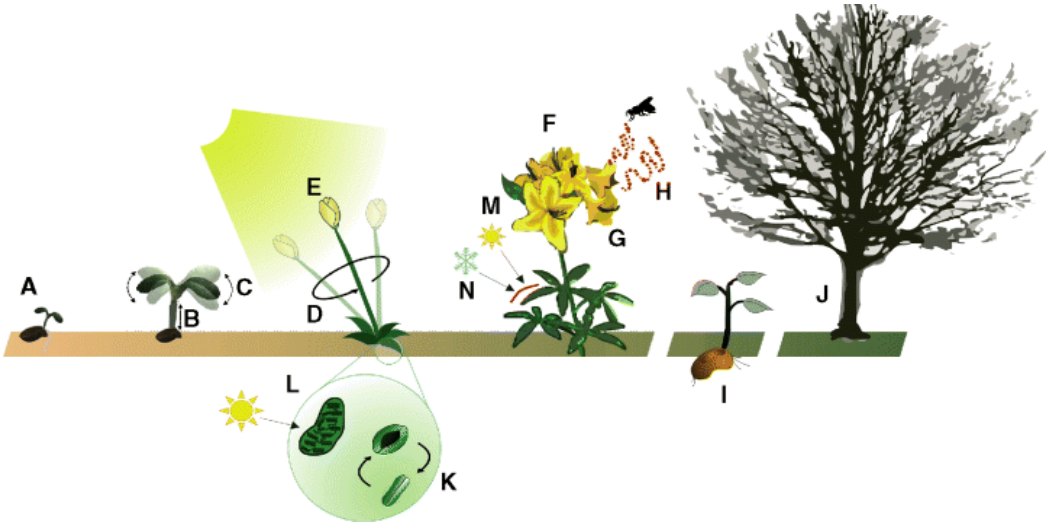


Figure 1.8 The regulatory role of the circadian system affects most aspects of the plant life-cycle. (A) Germination, (B) hypocotyl elongation, (C) leaf movements, (D) circumnutations, (E) shade avoidance, (F) flowering time, (G) flower opening, (H) scent production, (I) tuberisation, (J) winter dormancy, (K) stomatal opening, (L) photosynthesis, (M) photoprotection, and (N) protection from temperature extremes (Yakir et al., 2007).

Studies in *Arabidopsis thaliana* of clock gene mutants revealed that a core circadian oscillator feedback loop is necessary in plants to connect morning- and evening-phased loops (Pruneda-Paz and Kay, 2010). The two Myb-type transcription factors CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY) have peak oscillations around dawn (Yamashino et al., 2008), and reciprocally regulate another transcription factor, TIMING OF CAB (TOC1) (Alabadi et al., 2001). These two MYB TF(s) are morning expressed and TOC1 is an evening-phased pseudoresponse regulator (Alabadi et al., 2001; Strayer et al., 2000; Wang and Tobin, 1998). *CCA1* and *LHY* act as partially redundant genes (Mizoguchi et al., 2002; Salome and McClung, 2005) that directly repress TOC1 by binding to the Evening Element (EE) motif (Alabadi et al., 2001; Harmer et al., 2000) in the gene promoter. This EE motif (AAATATCT) is frequently overrepresented in clock-regulated gene promoters with a peak expression in the evening. Regulation via the EE motif is sufficient to drive the evening phased rhythms (Harmer et al., 2000; Harmer and Kay, 2005; Pruneda-Paz and Kay, 2010). The evening module also consists of the GATA motif which is associated with clock responsive genes (Manfield et al., 2007).

In legumes, homologs of the *Arabidopsis* circadian system have been studied in soybean (*Glycine max*), pea (*Pisum sativum*), common bean (*Phaseolus vulgaris*) and *Medicago truncatula*. The *Arabidopsis* clock homologs in legumes follow similar expression patterns, indicating that the legume plant clock is conserved and similar to that of *Arabidopsis* (Song et al., 2010). For example, functional characterisation of two pea clock genes *LATE BLOOMER1* (*LATE1*; an ortholog of *Arabidopsis GIGANTEA*, *GI*) and *DIE NEUTRALIS* (*DNE*; ortholog of *Arabidopsis ELF4*) in the mutants *late1* and *dne1* were found to be photoperiod insensitive (Hecht et al., 2007; Liew et al., 2009). Under continuous weak light and dark conditions, the pea *late1* mutants not only exhibit a stronger clock phenotype than the Arabidopsis *gi* mutants but also show arrhythmic expression of pea *TOC1*, *DNE* (*ELF4*) , and *MYB1* (*LHY* homolog) genes (Hecht et al., 2007).

Study of *CCA1* and *PRR5* (*Pseudo Response Regulator* 5) in the model legume *Lotus japonicus* found that the circadian rhythm in heterologous cells of *L. japonicus* is comparable to that of Arabidopsis core clock genes (Ueoka-Nakanishi et al., 2012). In soybean, the existence of diurnal oscillations and its connections to drought stress have been studied. The expression of several circadian genes such as *LCL1*-, *GmELF4*- and *PRR*-like genes were reduced in drought-stressed soybean with a phase advance of expression for the *GmTOC1*-like, *GmLUX*-like and *GmPRR7*-like genes (Marcolino-Gomes et al., 2014). Analyses of the upstream promoter region of these soybean genes also revealed the presence of cis-regulatory elements associated both with circadian clock and stress (Marcolino-Gomes et al., 2014).

1.9 Plant circadian clock in roots and nutrient acquisition

The molecular mechanisms of the circadian clock have been mostly studied in leaves or hypocotyls, with an emphasis on investigating the interactions of day-night cycles (de Montaigu et al., 2010; Nozue et al., 2007). In plants, the above ground shoot system perceives the light signal, and the underground root system is involved in water and nutrient acquisition.

Oscillatory rhythms in root growth rates under diel conditions (constant light and dark) have also been demonstrated in rice and Arabidopsis (Iijima and

Matsushita, 2011; Yazdanbakhsh et al., 2011). Different rhythmic properties (period and amplitude) of the root and shoot circadian clocks have been shown to have different responses to light, indicating an organ-specificity in plant circadian system (Bordage et al., 2016). For example, in Arabidopsis, the root circadian clock has been shown to be different than the shoot clock. The root not have an evening loop because transcription factors *CCA1* and *LHY*, inhibit gene expression in shoots but not in roots (James et al., 2008).

Apart from the various diel growth rhythms described in vascular plants (Walter et al. 2009), plant growth is regulated by environmental factors such as water and nutrient availability. Water and nutrients acquired by the roots are transported to the plant shoot and plant growth at the molecular level is modulated through a complex gene regulatory network whereby circadian clock acts at multiple levels (Farre, 2012).

NMR microimaging in *Arabidopsis* has been used to discover that root water content oscillates with a peak at dusk (Takase et al. 2011) under both constant light and darkness. This study showed that water transport in the root and the expression of two aquaporin genes *AtPIP1;2* and *AtPIP2;1* were under the regulation of circadian rhythm. The circadian rhythm affects N-assimilatory pathway such as glutamate, glutamine and other N-metabolites under the regulation of master TF clock CCA1 (Gutiérrez et al., 2008) and an expanding network of CCA1 oscillatory components (Figure 1.9, Kay, 2010). In *Arabidopsis* lateral roots (LR), overexpression of core clock components *LHY* and *CCA1*, represented by a single gene in *M. truncatula* and *G. max*; (Hecht et al., 2005), have been shown to oscillate in LR primordia and rephase during LR development (Voss et al., 2015). As such, investigation of the circadian clock in another root lateral organ, the legume nodule, is highly relevant for understanding how the rhizobium-legume symbiosis is regulated.

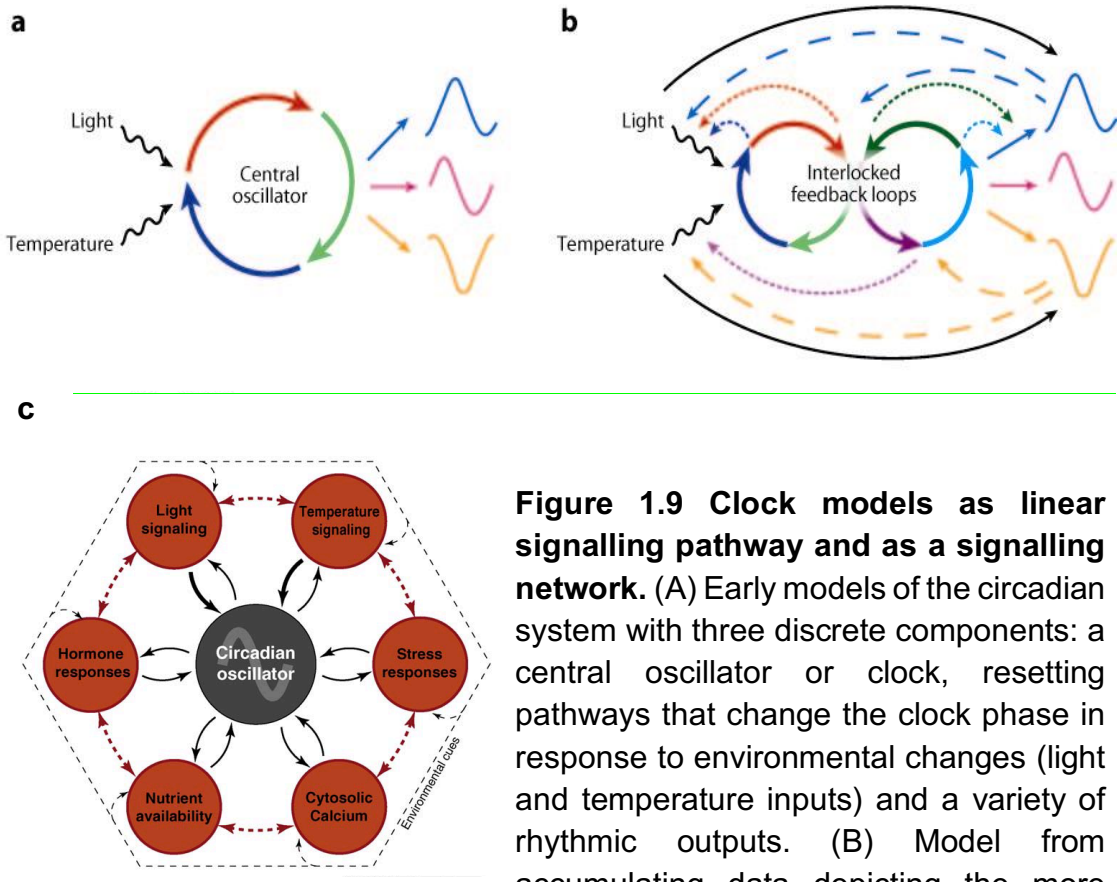


Figure 1.9 Clock models as linear signalling pathway and as a signalling network. (A) Early models of the circadian system with three discrete components: a central oscillator or clock, resetting pathways that change the clock phase in response to environmental changes (light and temperature inputs) and a variety of rhythmic outputs. (B) Model from accumulating data depicting the more complex network of the circadian system. The oscillator consists of multiple coupled feedback loops (solid coloured lines). Clock genes demonstrating multiple functions, acting within the oscillator and in clock input and output signalling pathways (dotted lines). Clock outputs can feed back to regulate clock components and input signalling pathways (dashed lines). Likewise, input pathways can regulate multiple clock genes and directly affect clock outputs (solid black lines) (Harmer, 2009). (C) The expansive regulatory network where the circadian oscillator and major plant signalling modules (red circles) are regulated in a reciprocal manner (black curved arrows, thick lines denote the higher importance as clock inputs). (Harmer, 2009; Michael and McClung, 2003; Pruneda-Paz and Kay, 2010)

1.10 Aims and Objectives

This work aimed to investigate the intersection of control of nodulation, signalling and root development. The new work presented adds to the complexity of understanding of RL symbiosis and suggests a new role for members of the large gene family encoding for nodule cysteine rich peptides (NCRs). Briefly the investigation can be outlined as:

Chapter 3: Root system architecture (RSA) for A17 and *sunn-1* plants in nitrogen depletion or repletion in +rhizobia/-rhizobia was analysed. We utilized the *sunn-1* mutant because of its hyper-nodulating phenotype and disrupted AON signals. Results from root phenotypic studies were integrated with the transcriptomic data of A17 and *sunn-1* mutant. Using a systems biology approach we identified set of genes responsive to N status and rhizobia, including NCRs, that could be implicated in the rhizobia/nitrogen regulation of root development.

Chapter 4: *De novo* motif discovery from the differentially expressed NCR genes was carried out. Whether the motifs were a feature of NCR promoters was analysed by testing if they were over-represented compared to other promoters of genes in the Medicago genome was carried out. This implicated the motifs to be feature of the NCR promoter genes. The motif landscape was evaluated, enabling smaller motifs to be found to occur as part of longer stretches of conserved sequences.

Chapter 5: An *in silico* TF search from the *Arabidopsis thaliana* database found TF(s) - CCA1, RVE1, ATHB15, ATHB16, AHL20 and AHL25 as putative motif binders. An orthology-based analysis was used to identify Medicago orthologs of these putative TF(s) and use of our A17-sunn transcriptomic data found them to b were found to be expressed genes in tissue that we hypothesised them to be active in. Quantitative PCR confirmed the perturbed level of mRNA expression in the homozygous lines identified from the Medicago mutant lines.

Chapter 6: Transcriptomic time series data of A17 nodulating roots over a period of 48 hours was performed and gene expression changes investigated. We hypothesised that NCRs have a role beyond acting as anti-microbial peptides, and if the over-representation of circadian binding sites on NCR

promoters was functional, that the NCRs would oscillate over time. Preliminary findings from the time series are presented.

Impact of this PhD thesis

A combination of promoter analyses, systems biology integration of phenotypic and transcriptomic data, including a time series, for the model legume *Medicago truncatula* has laid an exciting foundation to suggest the existence of a nodule circadian clock. This new data enable the investigation of root circadian links to legume-rhizobium symbioses.

Chapter 2: Materials and Methods

2.1 Plant Material

Medicago truncatula seeds cv. Jemalong line A17 (sequenced reference accession, denoted here as 'wild type') were obtained from the IGER seed bank (Aberystwyth, http://www.igergru.ibers.aber.ac.uk). Seeds of the A17 background with hypernodulating phenotype *sunn-1* mutant (super numeric nodules) (Schnabel et al., 2005) was kindly provided by Giles Oldroyd (John Innes Centre). 37 Medicago Tnt1 lines in the R108 background was procured from The Noble Research Institute, Oklahoma, US with the kind help from Jiangqi Wen. The Medicago lines that were found to be defective from the BLAST search all the ortholog genes were (Expressed ortholog lines: NF2784, NF17115, NF13921, NF1531, NF17813, NF2932, NF9265, NF3772, NF5835, NF13955, NF1297, NF5993, NF17550, NF20691, NF16461, NF17277, NF20547, NF3697, NF11111, NF1264, NF20437, NF0908, NF1084, NF1729, NF14210 NF11854) and (Non-expressed ortholog lines: NF8126, NF12227, NF7335, NF5999, NF14591, NF19358, NF17814, NF17813, NF5993, NF10395, NF19167, NF21293) Also see Table 5.1-1 and 5.1-2 with their corresponding genes and co-ordinates.

2.2 Plant Growth

2.2.1 Medicago seed extraction and scarification

Seeds were extracted from the pods by crushing on a corrugated rubber mat with a plasterer's hawk. Care was taken to protect the seeds from damage during pod-crushing and the seeds were collected by a pair of forceps into in a 50 ml Falcon tube. Chemical scarification with 3-5 ml of concentrated sulphuric acid was used to break the outer seed covering which appeared as 5-6 dark spots on each of the seeds. Excess H_2SO_4 was removed using a pipette then seeds were thoroughly rinsed (3-4 times) with sterile water.

2.2.2 Surface sterilisation and germination

Seeds were sterilised in bleach:water (7% sodium hypochlorite) with intermittent agitation for 3 min, then rinsed in sterile water for 4 times. Sterile seeds were soaked in the last wash of water for 4-5hours of imbibition. Seeds were then sown on 1.5% phyto-agar in sterile 12 cm square petri-plates (Greiner Bio-one) and sealed using 1.25 cm x 9.1 m microporous tape (3M™ MicroporeTM). Seeds were wrapped in aluminium foil and kept in the dark in a 4°C cold-room for optimal synchronisation of germination for 3 days. Seeds were then incubated inverted (to promote straight radicle emergence) in a growth chamber (MLR-351H, Sanyo, E&E Europe BV, Loughborough, UK) at 25°C under fluorescent lights (16/8 hr light/dark photoperiod) for 5 days.

2.2.3 Medicago seedlings growth on phyto-agar

Germinated Medicago of 6-7 seedlings were transferred on modified low N Fahräeus medium in between two layers of a growth pouch (CYG™ Germination Pouch, West St. Paul, MN, United States) overlaid on solid Fahräeus medium (Vincent, 1970). The modified N-free Fahräeus medium contained the following macronutrients (0.5 mM $MgSO_4.7H_2O$, 0.7 mM KH_2PO_4, 0.4 M $Na_2HPO_4.2H_2O$, 0.1 mM) and micronutrients (20 mM Ferric citrate, 8 mM $MnSO_4$, 4 mM $CuSO_4$, 7.34 mM $ZnSO_4$, 16 mM H_3BO_3, 4.13 mM Na_2MoO_4). The source of N as NH_4NO_3 was added as 0.1 mM for deplete and 5 mM for N-replete conditions.

The pH of media was adjusted to 6.5 using KOH and 1.5% (w:v) phyto-agar was added before autoclave sterilization. Then 0.075 µM (S)-trans-2-Amino-4-(2-aminoethoxy)-3-butenoic acid hydrochloride (AVG) and 1 mM $CaCl_2$ was added to the cooled (around 60°C) medium. The medium was then poured into sterile square Petri dishes Corning 500 cm^2 Square TC-Treated Culture Dish (Wilford Industrial Estate Ruddington Lane, Nottingham, UK); AVG is an ethylene inhibitor (Peters and Crist-Estes, 1989) and it is needed for enhancing nodulation by promoting infection events of the *S. meliloti* (Oldroyd et al., 2001; Penmetsa et al., 2003). To make the experiments comparable, AVG was added to the growth medium in all experiments for WT Medicago A17 and *sunn-1*. Plates were wrapped with black plastic bags (to cover the roots to

mimic soil conditions) and grown vertically in the growth chamber (25◦C, 16/8 hr light/dark) for 4 days.

For Medicago Tnt1 lines that had been genotyped, homozygous and WTS plants along with with R108 as wild-type background control were germinated and transferred to modified Fahräeus agar containing 5 mM of NH_4NO_3 as above. Tnt-1 lines for the *ATHB15* gene family were grown on MFM for 7 days and Tnt-1 lines for *CCA1* and *RVE1* genes were grown for 25 days. The difference in the duration of the growth is because *cca1* and *rve1* mutant lines took a longer time to properly develop roots; preliminary phenotypic data not shown.

2.2.4 Plant growth conditions for characterisation of Tnt-1 mutants

For Tnt-1 Medicago lines the germinated seedlings were transplanted onto a Medicago soil mixture of 1 part Levingtons F2s, 1 part Vermiculite plus Osmocote added at 3 gm per liter. Plants were grown in a glasshouse regulated to 22ºC with a 16/8 hr light/dark photoperiod. The pots were covered with plastic lids for 3-4 days to maintain high humidity. Leaf disc samples were collected for genotyping 10 days after transplantation. Plants were grown in glasshouse until maturity (typically 3 months). Watering was discontinued during the seed-setting stage. All the plants were then covered with transparent plastic bags for drying and harvest pf seedpods. Harvested seeds were kept in paper bags at room temperature in laboratory conditions.

2.3 Rhizobium treatments

2.3.1. Growth of rhizobia on solid TY media

Rhizobium strain *Sinorhizobium meliloti* 1021 was cultured on agar TY medium and $CaCl_2$ (Journet et al., 2006) composed of 5 g l^{-1} bacto-tryptone, 3 g l^{-1} yeast extract and 1.2% (w:v) bacto-agar into sterile deionised water. The medium was then autoclaved and 6 mM $CaCl_2$ was added to the cooled (around 60ºC) medium. Rhizobial liquid culture from a stock was streaked out onto the solid medium and incubated for 3-4 days at 28ºC.

2.3.2 Rhizobial liquid cultures for inoculation

Rhizobium inoculants was prepared by spot-inoculating rhizobia into liquid TY/Ca^{2+} medium incubated at 28ºC while shaking at 220 rpm to an OD_{600} of 1 as measured using a spectrophotometer. Cells were harvested by centrifugation (4000 rpm, 10 min, 4ºC) and re-suspended in 20 ml liquid modified Fahräeus medium (supplemented with the same concentration of NH_4NO_3 as the modified Fahräeus agar medium). The final OD_{600} value of resuspended rhizobium culture used for inoculation was 0.02-0.05. In all conditions, 4d old Medicago seedlings grown in modified Fahräeus media were inoculated evenly with 2 ml of *S. meliloti* 1021 solution (OD_{600} = 0.02-0.05). Mock inoculation (as control) was carried out by treating seedlings with 2 ml of sterile liquid modified Fahräeus medium without rhizobium.

2.3.3 Plant treatment and growth conditions for A17 and *sunn-1* microarray and A17 time-course experiment

For A17 and *sunn-1* microarray analysis, mock- and rhizobia- inoculated A17 and *sunn-1* plants were incubated in the growth chamber for 14 days after germination. At 14 days post inoculation (dpi), roots were treated with 2ml liquid modified Fahräeus medium supplemented with NH_4NO_3 (deplete or replete N concentration) per plate. Inoculated plants on plates were kept in the incubator for further 2 days. Hence, plants were incubated for 16 dpi for analysis of root system architecture. For the A17 time-course (48 hrs) experiment, rhizobium inoculation and transfer on to the modified deplete N Fahräeus agar medium (0.1 mM NH_4NO_3) were performed on the same day for the A17 germinated seedlings. Plates were sealed, wrapped in black plastic bags and kept in the growth chamber for 7 days (16/8 hr light/dark photoperiod). 7 dpi roots were treated with 1-2 ml liquid modified Fahräeus medium supplemented with NH_4NO_3 (deplete or replete N concentration) per plate then placed on fresh plates of Fahräeus agar medium supplemented with the same NH_4NO_3 concentration as the treatment. All the inoculated plants were grown in deplete or replete N conditions for 0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44 and 48 hours in the growth chamber. The whole root system with nodules was harvested in liquid N_2 and stored at -80°C for grinding.

For the timeseries, four biological replicates of Medicago A17 wild type seeds were scarified, germinated, grown on low N Fahräeus medium and inoculated with rhizobia as described above. However, the inoculation was performed on 4 days old seedlings with transfer to low N for growth on the same day. Then 7 dpi whole plants were transferred to fresh Fahräeus medium with replete N (5 mM) to collect the whole roots every 4 hours for a period of 48hours.

2.3.4 Probe specificity for the NCR genes on the NimbleGen microarray

Roche NimbleGen microarray probe design is based on the methodological strategy similar to WindowMasker program (www.repeatmasker.org). Based on this strategy, to avoid non-specific binding, Roche NimbleGen identified and excluded highly repetitive elements in the genome or transcript sets for probe design (Morgulis et al., 2006). In this thesis, the hybridisation specificity of the NimbleGen probes of NCR genes was analysed with a nucleotide BLAST method. Probe sequences for each of the 185 NCRs on the Nimblegen microarray, represented by 2-3 probes each (530 probes in total), were used to BLAST against *M. truncatula* 4.0 genome on Ensembl plants (plants.ensembl.org). BLAST results of 10 randomly selected NCR genes are given on Appendices Table 2 as an example.

2.4 Phenotyping of Medicago roots and nodules

2.4.1 Whole root architecture quantification

Plant roots were scanned using a Scanjet G2710 flatbed scanner (Hewlett-Packard) at highest resolution to image for all the plants in biological replicates (6 seedlings per plate). Roots of A17 and *sunn-1* plants for microarray experiments were scanned at 14 dpi. For A17 plants in time-course experiments phenotyping was carried out at 21 dpi after rhizobium inoculation (14 days after the plants were transferred to deplete or replete N conditions). The number of LRs and nodules (for rhizobia inoculated seedlings) were counted. Root architecture was measured using ImageJ (http://rsbweb.nih.gov/ij) (Schneider et al., 2012). The primary root (PR) length,

the number of lateral roots (LR number) and the length of every LR was measured. From this the following were calculated: total LR length, average LR length (LR length ave), total LR plus PR length (PR+LR tot) and lateral root density (the number of LRs per cm of PR).

2.4.2 Statistical analysis of root system architecture

Phenotype data was analysed using R with a Shapiro-Wilk test used to test data normality and a Bartlett test was used to test data variance. A pairwise Wilcox test was used to assess differences for significance using the Benjamini-Hochberg method. Data boxplots were generated using the R package ggplot2.

2.5 Genotyping primers

2.5.1 Primer designs for Tnt-1 Medicago lines screening

2.5.1.1 Tnt-1 and gene specific primer (GSP) design

Forward primers and reverse primers for Tnt-1 insert Medicago mutants, Tnt1-Fg and Tnt-1Rg, Tnt1-Fg1 and Tnt1-Rg1 (Table 2.5-1) were used for PCR screening. Tnt primer sequences were used as (Cheng et al., 2014). Primer3 (version 4.1.0) was used to design the gene specific primers (GSP) based on R108 genomic sequence at ~1000 bp away from the flanking sequence target (FST) site for each of the mutants. Lack of secondary structure formation was checked on Beacon designer (Premier Biosoft International). The primers were designed to match the properties of the Tnt-1 primers as; primer length of 22-24 bp, 35-45% GC content, with no G or C clusters longer than 4 and presence of either G or C at the 3' end of the primer. Gene specific primers were tested for amplification efficiency on A17 and R108 genotype DNA before use in genotyping. The sequence of all the tested GSPs used in the mutant genotyping are listed on table 2.5-1 and 2.5-2.

Table 2.5.1 List of *Tnt-1* primers and sequences

No.	Primers (Tnt-1)	Sequence 5'-3'
1.	Tnt1Fg	ACAGTGCTACCTCCTCTGGATG
	Tnt1Fg	CAGTGAACGAGCAGAACCTGTG
2.	Tnt1Fg1	CCTTGTTGGATTGGTAGCCAACT
	Tnt1Rg1	TGTAGCACCGAGATACGGTAAT

Table 2.5.2 List of GSPs used for genotyping and sequences

No.	Primers (GSPs)	Sequence 5'-3'
1.	CCA1-2-Fp2	CTCAAAACATGGCGGCTTAC
	CCA1-2-Rp1	AGTGGCTGAGATTGGTTGTG
2.	NF16461-Fp1	AATGAACGATTTTAGCAGCGG
	NF16461-Rp1	TTTGGCCGTATGCAAATGTAG
3.	RVE1-Fp1	GTTTTTGCTCTGATTTGTGTTGG
	RVE1-Rp2	GCAACTCCAATCATGGTCGTG
4.	NF1729-Fp1	AGTTTGGAAGATGGAAGTCTTGT
	NF1729-Rp1	GACTAAGCAGATTTGAGAATTGAG
5.	NF0908-Fp1	ATTGGTTCATGTTGATGGGAA
	NF0908-Rp1	AGATGACAATTGAATCGTGGC
6.	NF5993-Fp2	TAGGAGGACCCCGTGAAAATAG
	NF5993-Rp1	CTCTTCCCCTTCTTGTTGGAGA
7.	NF5999-Fp1	AGGGGCTAAATTGATGTTTTG
	NF5999-Rp1	GACAAGAGTTGGTGAAATGTG
8.	NF12227-Fp1	GTCAACAACAACACAGCTCAC
	NF12227-Rp1	CTAAGATGCATGTTTCCTCAGAG
9.	NF5835-Fp1	ACCCATGGAAAAAGAGTGGG
	NF5835-Rp1	AGCCAGCGATGTTACAAGAAG

2.6 Genomic DNA isolation

2.6.1 Plant genomic DNA extraction

The rapid 5% Chelex method using Chelex 100 chelating resin was used for genomic DNA extraction for Tnt-1 line genotyping (HwangBo et al., 2010).

Leaf discs of plant leaves were crushed in 120 µl Chelex solution at room temperature until a homogenized mixture was generated. Samples were vortexed and heated in Thermocycler at 99ºC for 10 mins then cooled down at room temperature and centrifuged at 4000g for 10 mins. 40µl of the supernatant was transferred to new tubes and a 1:10 dilutions were made for use in polymerase chain reaction (PCR).

2.6.2 PCR conditions and gel electrophoresis

For Medicago Tnt-1 line genotyping, PCR was performed in a thermocyler (Applied Biosystems, Invitrogen, Carlsbad, CA, USA). MyRed Taq DNA polymerase (Bioline) was used for all the PCR reactions (Tnt-1 and GSP). Each of 20 µl reaction mixture contained 4 µl 5x MyTaq red reaction buffer (Bioline Reagents Ltd, London, UK), 0.4 µl of each primer (forward and reverse), 0.2 µl MyTaq Red DNA Polymerase (500 Units) and 14 µl nuclease free water. PCR run was performed with a touch-down program (Cheng et al., 2014) as given on table 2.6.1.

The PCR product samples were then analysed using agarose gel electrophoresis. 1 % w/v agarose solution was prepared with 1xTAE buffer (Tris base, acetic acid and EDTA, prepared from a 50x stock solution). The agarose was melted in a microwave oven by heating and cooled prior to adding nucleic acid stain GelRed or SYBR safe. This mixture was poured into a gel casting tray (Thistle Scientific, Glasgow, UK) with a toothed plastic comb. The comb was removed after the gel had set and was placed in a gel running tank (Thistle Scientific) filled with 1xTAE buffer to cover the gel. PCR reactions (10 µl) were loaded into the wells. For product size comparison, known ladder 5 µl of 1Kb DNA HyperLadder I (Bioline) or 2 log DNA ladder (New England Biolabs, NEB) was loaded alongside the samples. The gel was run for ~50-60 mins at 90 V and 400 mA. The gel was viewed under ultraviolet light using a G-Box gel imaging system (Syngene International Limited, Biocon Limited, Bangalore, India).

Table 2.6.1 Touch down PCR program for genotyping Medicago Tnt1 lines

Phase 1	**Stage**	**Temperature (°C)**	**Time**
1	Denaturation	95	2 min
2	Denaturation	95	60 s
3	Annealing **x 5 cycles**	60	15-30 s
4	Elongation	72	60 s
Phase 2	**Stage x 5 cycles**	**Temperature**	**Time**
5	Denaturation	95	60 s
6	Annealing	57.5	15 s
7	Elongation	72	60 s
Phase 3	**Stage x 25 cycles**	**Temperature**	**Time**
8	Denaturation	95	60 s
9	Annealing	55	15 s
10	Elongation	72	60 s
11	Final extension	72	2.5-5 min
12	Hold	4	-

2.7 Real time quantitave PCR (qPCR) Primer design for Medicago mutants

Primer3 (version 4.1.0) was used to design primers for qPCR based on R108 genomic sequence within the last 500 bp of each gene and with a product size less than ~200 bp. Lack of secondary structure formation was checked on Beacon designer (Premier Biosoft International). In general the properties of primer pairs were: primer length of 22-24 bp, 35-45% GC content, with no G or C clusters longer than 4 and presence of either G or C at the 3' end of the primer. Primer efficiency or qPCR efficiency on A17 and R108 genotype cDNA was evaluated by using linear calibration curve and melting curve analysis. Linear calibration curves were obtained from serial dilutions (10^{-1} to 10^{-7}) of cDNAs in equal fraction. Linear regression analysis of Ct versus log(dilution) was used to derive the efficiency as, m = -(1/log E), where m is the slope of the regression line fit derived from standard curve and E is the efficiency of the primer (Schmittgen and Livak, 2008). Primers used as two reference genes for normalisation were *β-Tubulin* (Kakar et al., 2008) and *protodermal factor 2*, *MtPDF2* (Picard et al., 2013). The sequence of all the qPCR primers (with the

target genes) used to quantify the relative gene expression in homozygous, WTS and R108 Medicago plants are listed on table 2.7.1.

Table 2.7-1 List of qPCR primers used for relative mRNA quantification in the homozygous and WTS seedlings of Tnt1 *M. truncatula* lines

Medicago orthologous gene	**Primer sequence from 5'-3'**
β Tubulin	TTTGCTCCTCTTACATCCCGTG
Medtr7g089120	GCAGCACACATCATGTTTTTGG
MtPDF2	GTGTTTTGCTTCCGCCGTT
Medtr6g084690	CCAAATCTTGCTCCCTCATCTG
CCA1	CACAAAACAAAGAGAACGATGG
Medtr7g118330	ATGGCTCCTGATTTGCACAG
RVE1	CAGAGAGAAAAGTGGACAGATGAA
Medtr5g076960	AAAACTTCTGAGCATGACTTCG
ATHB15	CAGCACGAGACTTTTGGACACT
Medtr2g030130	GAAAACCACTTGGAAGCATTTC
ATHB15	GACTTGACATGCTTGAAACCAC
Medtr3g109800	GGATAAACAAATACCGCCTTGA
ATHB15	GGGGAACCATTGAGCTTATTTAC
Medtr4g058970	GCCATTGTCCAAAGTTGTAGTG

2.7.1 Quantitative PCR on Medicago mutants

Master mixes for qPCR were prepared using fluorescent dye SYBR Green I (Sigma-Aldrich, St Louis, MO) according to the manufacturer's protocol, with primer pairs listed in Table 2.7.1. A cycle of 94°C for 15 s, 60°C cycles of 94°C (15 s) and 60°C (30 s) was used for amplification in a LightCycler 480 system (Roche). *CCA1*, *RVE1* and *ATHB15* were normalised to internal control reference gene expressions (*β-Tubulin*, *Medtr7g089120*, and *MtPDF2 Medtr6g084690*). The expression of the reference/control gene *β-Tubulin* was checked in the transcriptomic time series data. Relative fold change was calculated by using mean C_T of the test genes and internal reference genes as follows: $E^{-\Delta Ct}$ of test genes/ $E^{-\Delta Ct}$ of reference genes, where ΔCt of test genes is given by ΔCt for test genes (calibrator-test gene) and ΔCt of reference genes (calibrator-test gene).

2.8 RNA isolation from whole root samples and cDNA synthesis

2.8.1 Extraction of RNA for expression profiling and qPCR

For microarray experiments, whole roots from four independent biological replicates were harvested and collected in aluminum foil bags into liquid N_2 for flash freezing. Root samples were stored for RNA extraction at -80°C (for A17 time-course experiments) or crushed in liquid N_2 and RNA isolated immediately (for real time qPCR of Tnt-1 Medicago mutants). The Monarch Total RNA miniprep kit from New England Biolabs (NEB) was used for all RNA extraction following the manufacturer's instructions. Frozen roots were ground completely in liquid N_2 in a chilled mortar and pestle then 400 ul of diluted protection buffer (from Monarch Total RNA miniprep kit, New England Biolabs, NEB) immediately added and mixed by pipette. All the extraction steps after this point were performed at room temperature following the manufacturer's instructions and RNA was eluted in nuclease free dH_2O and stored at -80°C for future use.

2.8.2 RNA integrity check and testing for gDNA contamination

RNA quality was checked by analysing the RNA samples mixed with 0.5% Orange-T dye using gel electrophoresis. To verify the lack of gDNA contamination from the RNA samples, a PCR of 35 cycles was performed using *β-Tubulin* as reference gene primer. The samples were then analysed using gel electrophoresis to compare with PCR products from a positive control reaction that used gDNA as template.

2.8.3 First strand cDNA synthesis for qPCR of mutant screening

An optimised mix of ProtoScript® II First Strand cDNA Synthesis Kit (NEB, US) was used for first strand cDNA synthesis of the clock (*CCA1, RVE1*) and *ATHB15* line RNA samples. Following the manufacturer's protocol, 200 ng of RNA template was mixed with a master-mix that contained Oligo $d(T)_{23}$, ProtoScript II Reaction Mix (2X), ProtoScript II Enzyme Mix (10X) and nuclease-free water. The mixture was incubated at 42°C for 1 hour then the enzyme heat-

inactivated by heating at 80°C for 5 minutes. The synthesised cDNA samples were flash-frozen in liquid N_2 and stored at -20°C.

2.8.4 cDNA synthesis and amplification for microarray timeseries

For microarray experiments, double-strand cDNA synthesis and amplification was carried out using the Ovation Pico WTA System (NuGEN Technologies Inc., San Carlos, CA, USA) following the manufacturer's instructions. The cDNA was then purified using Qiagen Qiaquick PCR purification kit following the manufacturer's instructions. The cDNA was quantified using a Nanodrop set to measure the quantity and purity of the cDNA.

2.9 Promoter analyses of A17 and *sunn-1* microarray data

2.9.1 Motif search

The [-200,-1], [-500, -1] and [-1000, -1] upstream regions of all the 62319 genes in Mt4.0 genome version were retrieved from the J. Craig Venter Institute http://www.jcvi.org/medicago/. The upstream sequence was used for the search with a parameter setting of motif minimum width 6 and maximum width of 12 bp. Promoter motif analysis for groups of co-regulated genes from hierarchical clustering was carried out by using MEME-LaB (Brown et al., 2013) and MEME suite (Bailey et al., 2015). *de novo* motif discovery runs were performed on either strand of unaligned sequence within the set of co-regulated genes and the promoter region was restricted to 500 bp upstream of the transcriptional start site, TSS for each open reading frame, ORF as most of the regulatory binding sites fall into these regions (Spellman et al., 1998). To find strongly supported motifs the maximum number of motif outputs were restricted to 3 per analysis with the option of motif sites distribution set as zero or one occurrence per sequence.

2.9.2 Investigation of motif structure in Nodule Cysteine Rich secreted peptides (NCR) promoters

For each set of promoters, position weight matrix (PWM) scanning was performed using FIMO (Find individual motif occurrence) to determine the number of motif sites. As test data input, promoter subsets of 4 specified

backgrounds were grouped together. The number of promoters in each set to be used as a training sequence data set was normalised to 185 (as 185 NCR genes were expressed based on analysis of the A17-*sunn-1* microarray data). The background gene dataset was randomly sorted to form a non-overlapping set of 185 genes from the remainder of the *M. truncatula* genome that included non-expressed NCRs and all other non NCR genes The promoter sequences chosen to serve as background testing datasets were 1. expressed NCRs; 2. Non-expressed NCRs; 3. expressed non-NCRs; 4. non-expressed and non-NCRs. FIMO runs were performed three times for each motif against each 4 backgrounds. FIMO searched for MEME generated motifs against the 4 background datasets that consisted of Mt4.0v1 upstream sequences with a p-value < 1e-4 by scanning both the strands (Grant et al., 2011). After the scan, the mean motif hits were calculated and plotted using the R package ggplot2.

2.9.3 Multiple sequence alignment of NCR promoters

The upstream sequences of the Medicago genes retrieved from the Medicago genome database J. Craig Venter Institute http://www.jcvi.org/medicago/ were used for sequence alignment. A maximum of 500 nucleotides sequences were aligned using MAFFT (Multiple Alignment using Fast Fourier Transform) tools at EBI (European Bioinformatics Institute) (Katoh and Standley, 2013). Eight differentially expressed NCR genes were selected as representatives based on the presence of identified motifs in their promoters and motif 3 and motif 6 were positioned as two anchoring points. To give an optimal alignment for representation of motif conservation, we ran algorithms with a parameter setting of gap lowest open penalty 1 allowed in MAFFT, gap extension 0.5 and an iteration of 100 runs.

2.9.4 Alignment view and editing

JALVIEW version 10.9. (Waterhouse et al., 2009) was used to view the alignment generated by the MAFFT algorithm run. The processed files generated from MAFFT algorithm run were loaded as input sequences as .clustalw format. The alignment view was then manually edited for visualisation.

The same alignment parameters and file format were also used to view the alignment in Genious (version 11.0.2).

2.10 Transcription factor search

2.10.1 Tomtom motif comparison

Tomtom, implemented in MEME Suite (Gupta et al., 2007), was used to quantify the similarity of *de novo* identified putative motifs to known motifs. Tomtom was used to search for query motif against the database of known motifs, taking all possible offsets and orientation of the DNA strands. Each 500 bp upstream region was probed for the conserved binding sites. PWMs were created using 10 bp rolling frames over the most conserved regions as evident from the multiple sequence alignment; a rolling frame here refers to 10 bp stretch of nucleotide which all overlap by 5 bp to ensure the whole region is covered twice for a Tomtom run. Tomtom was performed against the following 4 databases: JASPAR plants 2016, *Arabidopsis* PBM db from (Franco-Zorrilla et al., 2014) and the CIS-BP database from (Weirauch et al., 2014) for *Arabidopsis* (CISBP-At) and *Medicago* (CISBP-Mt).

2.10.2 Orthology based analysis for TF(s) in Medicago

A search for TFs was carried out in the Medicago genome by performing BLAST searches to compare each the amino acid sequences of *Arabidopsis* and Medicago TFs. A FASTA file of amino acid sequences for the genes of interest (*Arabidopsis* as query, Medicago as subject) in each genome was made in a BLAST database format. BLASTP (Altschul et al., 1990) was performed between each set of proteins using the tabular output setting and user specified significance threshold of 0.0001, The best hits from the BLAST comparison were visually assessed for conservation based on the significance E-values (~10-order magnitude difference in result hits) criteria.

2.11 Identification of Medicago mutants defective in orthologous genes

2.11.1 BLAST search against the R108 genome sequence

Unspliced gene-transcript sequences of A17 Medicago (including introns and UTRs) with co-ordinates were retrieved from JCVI. The A17 nucleotide sequences of all the ortholog genes were then indexed 200bp upstream and 200bp downstream from the open reading frame (ORF) using SAMtools (Li et al., 2009b) in Bash. The trimmed A17 sequences were then used for BLAST search against the R108 genome (v1.0) http://www.medicagohapmap.org/tools/r108_blastform from the Medicago HapMap project that outputs an alignment score and R108 nucleotide co-ordinates . As R108 released sequences are in scaffolds, the R108 co-ordinates from the BLAST results could then be used for sequence extraction from the genome. R108 sequences were also kindly provided by Dr. Jiangqi Wen, Noble Research Foundation, Oklahoma, US.

2.11.2 Tnt1 Retrotransposon inserted Medicago mutants identification

A database of Medicago mutants (https://medicago-mutant.noble.org/mutant/) from Noble Research Institute, Oklahoma, US) containing Tnt1 (Tadege et al., 2008) and FNB (Fast Neutron Bombardment) mutant population (Sun et al., 2018) in the R108 background was queried to identify mutant lines. The retrieved R108 sequences of each Medicago ortholog genes along with 200 bp upstream and downstream of the ORF region (as described in section 2.11.1) were used in a nucleotide BLAST (blastn) search against the population of high confidence Tnt-1 FSTs using default parameter settings of E-value cut-off 10^{-6}. Medicago Tnt-1 lines were selected based on their E-values and % identity > 95. Other hits obtained on region outside of the queried co-coordinate sites were not considered as potential mutant lines defective in the orthologous genes.

2.11.3 FNB Medicago mutants identification

The BLAST search was extended to identify Fast Neutron Bombardment induced Medicago lines (Sun et al., 2018). The same query co-ordinates and sequences of R108 Medicago ortholog genes (as described in section 2.11.1 and 2.11.2) were used against the database of FNB-CNV (Copy number variation) sequences with default parameter settings. However, FNB Blast search resulted in identification of most of the hits outside the queried region of Medicago orthologous genes. Few generated hits Medicago FNB lines with low alignment identity (< 80%) and high E-values were not significant. Our queried gene sequence ranged from ~1000 to ~8600bp depending on the Medicago orthologous gene length. Hence, the short probeset sequence of FNB resulted in relatively high E-values in the BLAST search, because significant E-values calculation takes the query length into account. High E-values with a shorter sequence alignment have a higher probability of occurring in the database purely by chance. Moreover, with current FNB tools utilising reference genome-based tools it is hard to find the deletion region if the deletion occurs in the gap regions of the reference genome (Sun et al., 2018). Hence, FNB Medicago lines were not considered for further investigation.

2.12 Microarray hybridisation experiment for time series of WT Medicago A17

An oligonucleotide probe microarray of 12x135k from Roche-Nimblegen (Roche Applied Science, Penzberg, Upper Bavaria, Germany) was used to quantify transcript expression of Medicago mRNA. This platform used a custom design (OID36783) for the *M. truncatula* genome; each of 62,319 genes was measured with 2-3 unique 60mer oligonucleotide probes (total of 14,6171 probes) per gene.

Samples were randomised before being processed. The Nimblegen one-color DNA labeling kit was used following the manufacturer's protocol. Purified and amplified cDNA (1 µg) was labeled with Cy3 fluorescent dye for single channel microarray analysis. The labeled cDNA was quantified using Nanodrop set to measure cDNA quantities then 4 µg (recommended for

eukaryote cDNA) of the labeled cDNA was hybridized at 42°C for 16-20 hours using the Nimblegen hybridisation kit and equipment.

After washing and air-drying (using a microarray dryer) the slides were scanned on a Nimblegen MS 200 microarray scanner using the Nimblegen MS 200 Data Collection Software. Cy3-labeled signal was excited using the green laser (for one colour array, 532 nm wavelength) to measure the hybridised probe fluorescence intensities. Parameter control was set following the Nimblegen user's manual with recommended resolution scan of 12x135k at 2 µm. Scan, auto or manual alignment were performed as per Nimblegen user's manual using Nimblescan software. Feature extraction and data collection as .XY signal files (raw probe level data by coordinates on the microarray) was carried out by supplying the array coordinates (from the array design file, .ndf).

2.12.1 Normalisation and quality control microarray experiments

Microarray analysis was performed in the R statistical software environment. The pdInfoBuilder package and design file (.ndf) from the Nimblegen manufacturer was used to build an annotation package associated with the array design. The Oligo package (Carvalho and Irizarry, 2010) was used to import raw XYS files into the R session. For background adjustment, Robust Multi-Array Averaging (RMA) (Bolstad et al., 2003; Irizarry et al., 2003) was performed for all the slides using the Oligo package for R with quantile normalisation and summarisation (to gene level by median polish).

2.12.2 Plotting expression of the NCR genes over time

Normalised and background-corrected gene expression data were imported into R. NCR gene expression data for 48hours time-point in deplete and replete N conditions was extracted and plotted using the ggplot package in R. Promoters of the NCR genes were scanned for the presence of clock motifs (CCA1/RVE1) using FIMO as described in section 2.9.2 to sub-group the NCR promoter genes based on the presence of one or more CCA1/RVE1 motifs.

Chapter 3: Systems biology approach using root phenotyping and transcriptomic profiling identifies gene clusters responsive to nitrogen influx and rhizobial inoculation

3.1 Introduction

Establishment of legume symbiotic associations with soil-borne rhizobium bacteria regulate a number of developmental and metabolic changes in host plants and bacteria. While the root system of legume host plants form nodules to accommodate rhizobia, the bacteria in turn supply nitrogen to the plants through atmospheric N-fixation. Such coordinated processes of symbiotic nodule formation and N-fixation depend on the environment of soil N status.

Both lateral roots and nodules are lateral root organs that govern root system architecture development. Genetic links between the control of nodulation and other nitrogen-responsive root development pathways, including lateral root development have been identified (Huault et al., 2014) . However, there are very few studies that characterise root development during nodulation at the gene regulatory level.

To study the RSA of nodulating plants and molecular cross-talk between rhizobium and nitrogen responses in legumes, we utilised wild-type A17 and the hypernodulating *sunn-1* mutant that is impaired in correct AON signal controlling nodule numbers (Kassaw et al., 2015). Using a systems biology approach, here root phenotypic studies with analysis gene expression have been integrated to address the interaction of rhizobium and nitrogen during nodulation. Medicago wild-type A17 and *sunn-1* mutant were grown in deplete N (0.1 mM), inoculated with rhizobia or mock (- rhizobia) and at 14dpi transferred to N depletion or repletion. In this chapter, investigation of RSA measured at 16 dpi are presented (section 3.1.1) with microarray analyses (of whole root samples at 14dpi and treatment with 6 hours of N depletion or repletion, section 3.1.2). Three major regulated gene clusters and candidate genes strongly responding to combinations of rhizobium or nitrogen are discussed.

3.1.1 Response of root system architecture in A17 and *sunn-1* during interaction of rhizobium and N

Root architecture changes of *M. truncatula* were investigated in roots of A17 and *sunn-1* growing in deplete (0.1 mM) or replete (5 mM) N concentrations in absence or presence of *S. meliloti* (Fig 3.1 A). The deplete (0.1 mM) and replete (5 mM) concentrations of N were selected based on the N concentrations used at similar experiments on *Medicago* and *Arabidopsis* (Gifford, Unpublished results) and previous published studies (Jeudy et al., 2010). Two biological replicates (n=6 plants per plate) of A17 and *sunn-1* grown in deplete condition were included as mock (-rhizobia) or inoculated with *S. meliloti* at 14 days of growth after germination. Plants were then transferred to deplete or replete N conditions and left to grow for 2 more days before root attributes were measured (Figure 3.1 A).

In the presence of rhizobium, A17 nodule formation at replete/high N is inhibited (Figure 3.1.1 A, B). The primary root (PR) was significantly shorter compared to mock non-inoculated plants, irrespective of N status (Figure 3.1.1 C). Of note is that the shorter PR of rhizobia-inoculated plants was also found during replete N (Figure 3.1.1 C) when nodulation is usually inhibited. This suggests a close co-ordination at the expense of nodule formation and root development driven by external N status. Moreover, rhizobia inoculation and nitrogen treatments seem to be additive in their effects on PR length, as PR in the rhizobia-and-replete N condition was shorter than in any other condition studied (Figure 3.1.1 C). Between replete/deplete N and rhizobia/mock-inoculated conditions there was no significant difference in lateral root (LR) number or length for A17 (Figure 3.1.1 D–E; Table 3.1.1).

As described above, compared to rhizobia-inoculated plants, mock-inoculated plants had a longer PR (Figure 3.1.1 C, 3.1.1 H). In contrast *sunn-1* mutants had a significantly longer PR than A17 on replete N, either with or without rhizobia inoculation (Figure 3.1.1 H; Table 3.1.1-1). As reported in other previous findings (Jin et al., 2012), hyper-nodulant *sunn-1* developed significantly more number of nodules than A17 wild-type, even on replete N level (5 mM NH_4NO_3) when wild-type showed inhibition in nodulation (Figure 3.1.1 F, G; Table 3.1.1).

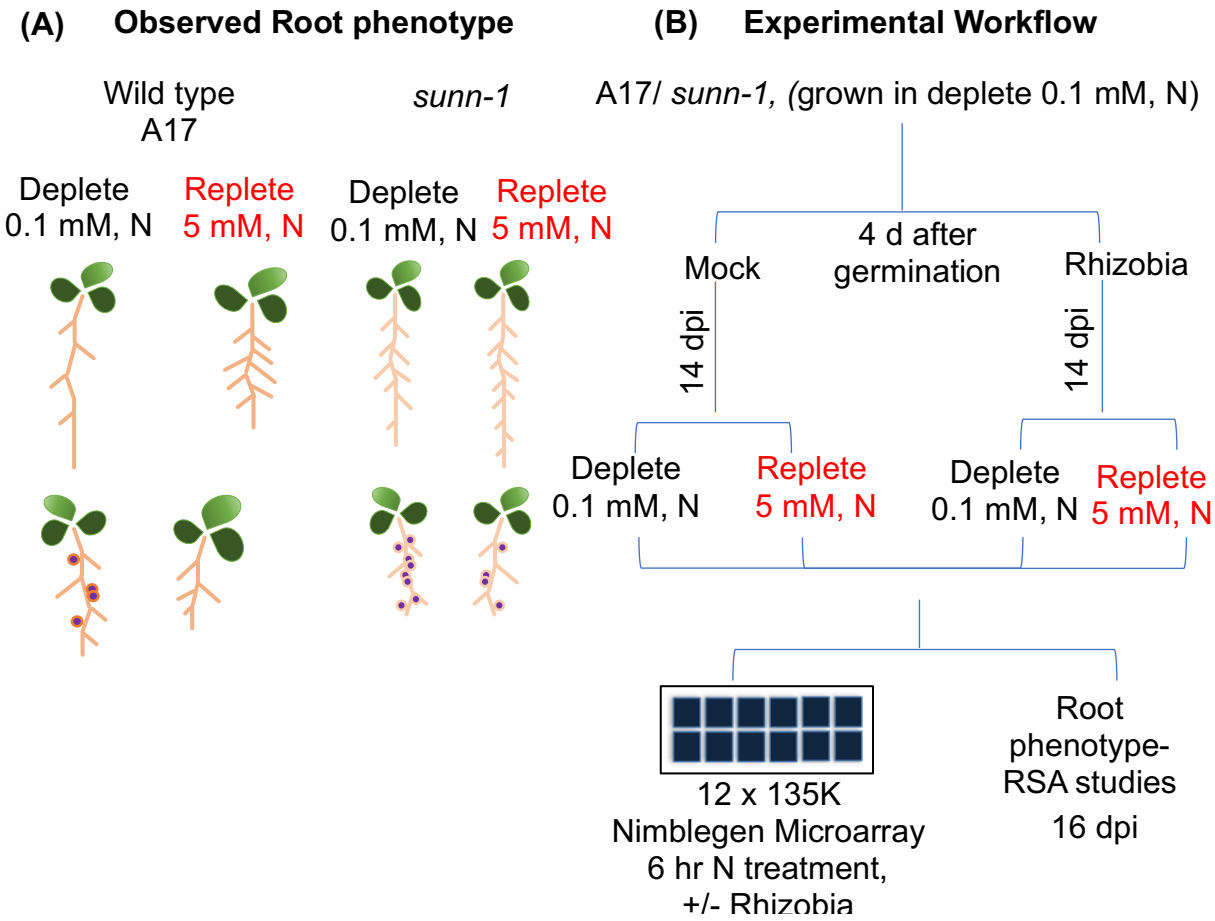


Fig 3.1: Systems biology approach to investigate effect of N status and Rhizobium during nodulation. All the plants were either mock or rhizobium inoculated on 4d old during N depletion condition. 14dpi all the plants were transferred to either 0.1 mM (NH_4NO_3) or 5 mM (NH_4NO_3) containing Fahraeus media.

(A) Root system architecture (RSA) analysis at 16 dpi –Root phenotype observed in wild-type A17 and *sunn-1* mutant grown in deplete N (0.1 mM), inoculated with *S.meliloti* or uninoculated (mock) and transferred to deplete (0.1 mM) or replete N (5 mM) conditions. Root architecture studies demonstrating nodulation in *sunn-1* mutant during N repletion as compared to A17 wild-type when nodulation is not enabled. Shorter primary root development is seen in both A17 and *sunn-1* when rhizobium is present as compared to its absence. More number of LRs are seen during N repletion in absence of rhizobium.

(B) Overview of the experimental design for microarray experiments after transferring A17 and *sunn-1* to treatment with N depletion or repletion for 6 hours.

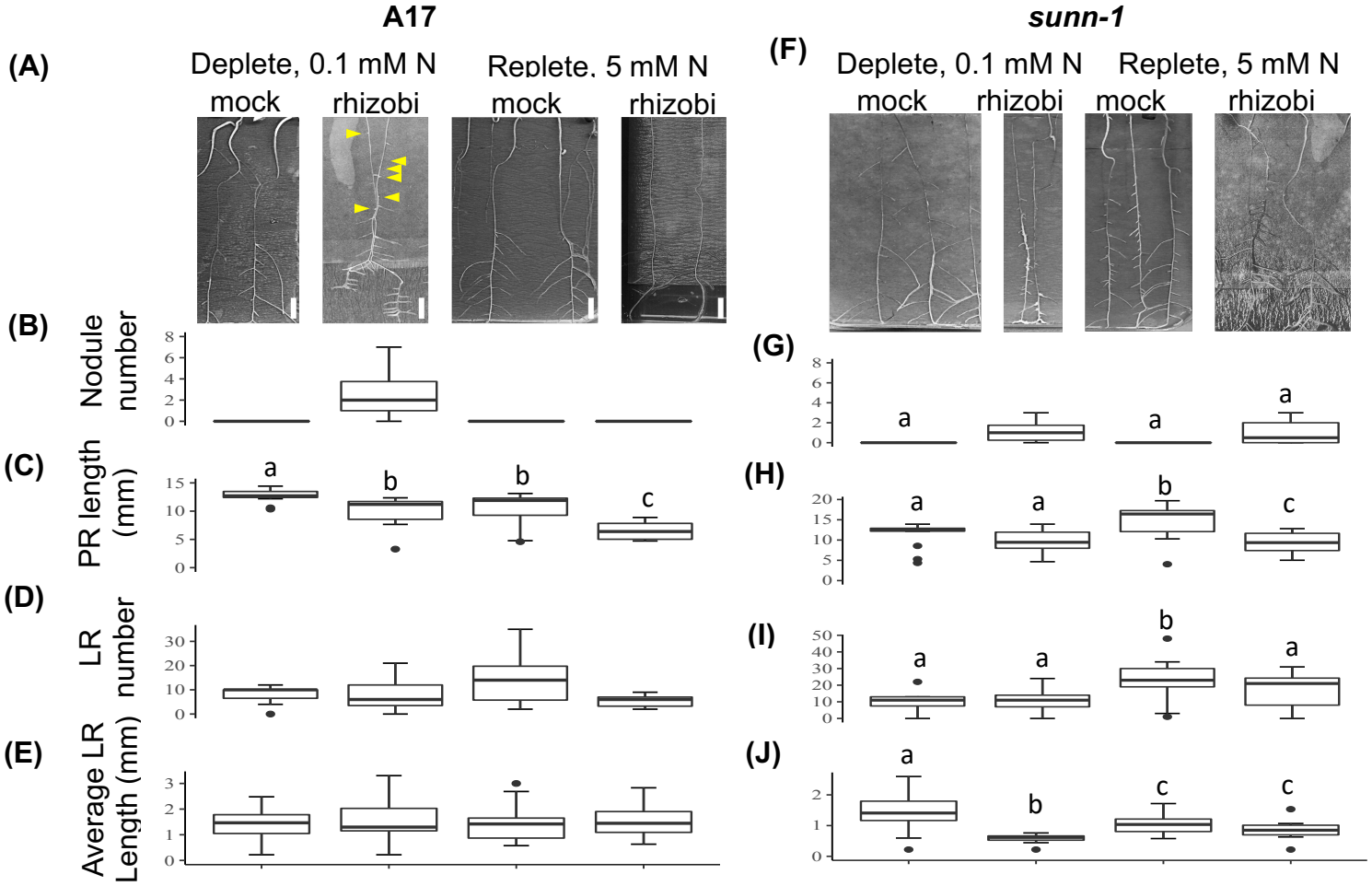


Fig 3.1.1: Root architecture in A17 and *sunn-1* plants in a nitrogen/rhizobia inoculation treatment space. (A) Images of A17 seedlings that were mock or rhizobia-inoculated then grown in N-deplete (0.1 mM NH_4NO_3) or N-replete (5 mM NH_4NO_3) conditions. (B-E) Root attributes in A17; (B) number of nodules (C) primary root length (mm), (D) number of LRs, (E) average LR length (mm); data are represented as mean ±SD; letters denote statistically different values, a-b or b-c $P < 0.05$; a-c $P < 0.01$; scale bar=1cm. (F) Images of *sunn-1* seedlings that were mock or rhizobia-inoculated then grown in N-deplete or N-replete conditions. (G) number of nodules (H) primary root length, (I) number of LRs, (J) average LR length; data are represented as mean ±SD; letters denote statistically different values, a-b or b-c $P < 0.05$; a-c $P < 0.01$; scale bar=1cm. See table 3.1.1-1

Table 3.1.1 RSA analysis for wild-type A17 *M. truncatula* and *sunn-1* mutant. Adjusted p-values from pairwise Wilcoxon rank sum test with Holm's method for multiple testing; Asterisk * indicates P-value < 0.05.

Root attributes	**Mock deplete (0.1 mM)**		**Rhizobia deplete (0.1 mM)**		**Mock replete (5 mM)**	
PR Length	**A17**	***sunn-1***	**A17**	***sunn-1***	**A17**	***sunn-1***
Rhiz 0.1 mM	**0.0037***	0.14844	-	0	-	-
Mock 5 mM	**0.0201***	0.25926	0.4134	**0.01567***	-	-
Rhiz 5 mM	**0.0000***	0.05863	0.0056	0.83275	0.0162	**0.01363***
LR number						
Rhiz 0.1 mM	1.0000	**0.023***	-	-	-	-
Mock 5 mM	0.5170	0.757	0.4350	-	-	**0.04***
Rhiz 5 mM	0.0960	0.128	1.0000	0.191	0.1310	0.414
LR ave						
length	1.0000	**0.00577***	-	-	-	-
Rhiz 0.1 mM	1.0000	0.08191	1.0000	**0.00219***	-	-
Mock 5 mM	1.0000	**0.01955***	1.0000	**0.0145***	**0.0322***	0.27728
Rhiz 5 mM						
PR+LR total			-	-		
length	0.4006	**0.0401***	0.4006	0.0144	-	-
Rhiz 0.1 mM	0.4006	0.1587	0.3386	0.1587	-	-
Mock 5 mM	0.0032	0.3753			**0.0322***	**0.0478***
Rhiz 5 mM						
Nodule no.			-	-		
Rhiz 0.1 mM	0.0001	-	0.0001	-	-	-
Mock 5 mM	-	**0.00017***	0.0001	**0.03088***	-	**0.00025***
Rhiz 5 mM	-	**0.00569***			-	**0.00709***

At N depletion, both A17 and *sunn-1* plants had similar numbers of lateral roots (Figure 3.1.1 D, I). But on N repletion, *sunn-1* mutants had significantly more LRs than A17 irrespective of rhizobium presence or absence (Figure 3.1.1 I). *sunn-1* mutants that were rhizobia-inoculated had shorter LRs than uninoculated ones only in deplete N conditions, (Figure 3.1.1 I), possibly as a resource balancing response because the plants were investing in increased nodulation.Rhizobia-inoculated *sunn-1* plants also had shorter LRs than rhizobia-inoculated A17 plants, both in N-replete and N-deplete conditions (Figure 3.1.1 D, I).

3.1.2 Gene expression profiling identifies gene clusters responsive to rhizobia and N-status

Phenotypic studies from RSA in both WT and the *sunn-1* mutant overall demonstrated shorter PR and LR development when rhizobium is present. But strong response of shorter PR development was seen in N repletion condition than in N depletion in both WT and the *sunn-1.* High N environment even in presence of rhizobium generally inhibits nodulation in WT Medicago plants. However, *sunn-1* mutants with shorter LRs in such environment (+rhizobia) still showed nodule formation. Such N resource and RSA balance suggested a complex regulation resulting from combinatorial effects of rhizobia and N. Hence, to investigate the transcriptomic changes and regulatory pathway a microarray analysis was used to identify underlying gene expression in response to rhizobia and N. Previous findings have shown that some of the regulatory pathways for LR and AON that governs nodule numbers are shared (Kuppusamy et al., 2009; Mathesius, 2003). It is also known that the *sunn-1* mutant exhibits different responses depending on N status due to defective AON signal (Jin et al., 2012). Thus, our hypernodulating *sunn-1* mutant served as an ideal background to compare the gene expression changes with wild type A17.

Microarray analysis was carried out from root samples of A17 and *sunn-1* plants that grew in varying concentration of N (source- NH_4NO_3) in *S. meliloti* inoculated or un-inoculated conditions (Fig 3.1 B) in the Gifford lab by a previous PhD student (Lagunas et al., (2019) Molecular Plant, In Review). All

the plants that were mock or rhizobia inoculated (14dpi) were treated with N (0.1 mM or 5 mM) for 6 hours before the microarray experiment. Within this thesis the data was re-analysed to focus on analysing genes with the up-to-date Mt4.0 annotation. Differential gene expression analysis identified a set of 6910 significantly regulated genes. Silhouette statistics in MatLab were used to classify and distinguish major gene expression patterns by hierarchical clustering. 11 clusters with 3 predominant patterns (clusters 4, 6 and 9) were found (Figure 3.1.2 A). These three major clusters represented 91% DEGs from a total of 6910 DEGs in *M. truncatula* 4.0 genome.

Cluster 4 (1283 DEG) consisted mainly of genes encoding for transmembrane proteins, receptor like kinases, cysteine rich peptides, disease and defensin related proteins. Gene family of nodule cysteine rich secreted peptides, NCRs, 159 in number from this cluster were found to be differentially expressed (Figure 3.1.2 B). Cluster 4 (Figure 3.1.2 B) showed prominent upregulation in both rhizobium inoculated A17 and *sunn-1*. However, expression was found to be more strongly induced in response to rhizobia in *sunn-1* than in A17, independent of the N treatment.

Major genes in this rhizobia-enhanced cluster include important genes encoding for TF(s) protein such as nodule inception protein (*NIN*, *Medtr5g099060*) (Vernie et al., 2015), nodulation signalling pathway (*MtNSP2*, *Medtr3g072710*) (Kalo et al., 2005), LysM receptor-like kinase (*LYK10*, *Medtr5g033490*) (Larrainzar et al., 2015), leghaemoglobins (10 genes), nodulins (31 genes out of which 2 genes were early nodulins and 25 genes were late nodulins), glycine rich proteins (19 genes), calcium binding proteins (5 genes), auxin related proteins (8 genes) and GRAS (4 genes). N-metabolism related (4 genes) out of which important genes include ammonium transporter 1 protein (*Medtr7g098930*), asparagine synthetase (*Medtr3g087220*) and N-acetyl-glutamate synthase (*Medtr5g015300*).

The fact that the majority of the genes in this cluster are involved in the nodulation pathway could indicate rhizobia-induction effect. Stronger induction of such nodulation genes in *sunn-1* than in wild-type support the effect of the molecular alteration underlying the hypernodulation root phenotype with *sunn-1* still able to nodulate at N repletion. Other regulatory genes in cluster 4 include

genes such as LRR receptor like kinases (16 genes), MYB (11 genes), and MADS box transcription factors (3 genes), zinc finger proteins (19 genes), and members of the F-box protein family (17 genes) as well as transport genes including peptide transporters (10 genes) and peptide/nitrate transporters (6 genes), amino acid transporter (6 genes). Other co-expressed genes involved in enzymatic pathway include sugar and starch metabolism related genes (26 genes), C related exchange and transport proteins (22 genes)

Genes in A17 and *sunn-1* exhibited opposite responses in cluster 6 (with 3137 DEGs) during N repletion and rhizobium inoculation. Cluster 6 consisted of annotated genes such as kinase (276 genes), transport (142 genes), disease resistance proteins (77 genes), myb transcription factor (23 genes), redox (26 genes), calcium (28 genes), nitrogen (35 genes) and UDP (37 genes), (Figure 3.1.2 B). A17 genes showed stronger induction than *sunn-1* in response to 6 hour of N treatment when rhizobium is present. Most of the Nitrogen-related genes were mainly found in this cluster, including nitrate transporters (4 genes), peptide/nitrate transporters (8 genes), nitrogen fixation proteins (2 genes), and also ammonium transporters (2 genes out of 12 ammonium transporters in the whole genome). Nitrate inducible GATA transcription factor (8 genes) were also found in this cluster (Zhang et al., 2015). All the P-values for gene enrichment (which have been represented as wordclouds in Figure 3.1.2B) in each of the clusters are listed in Appendices Table 1.

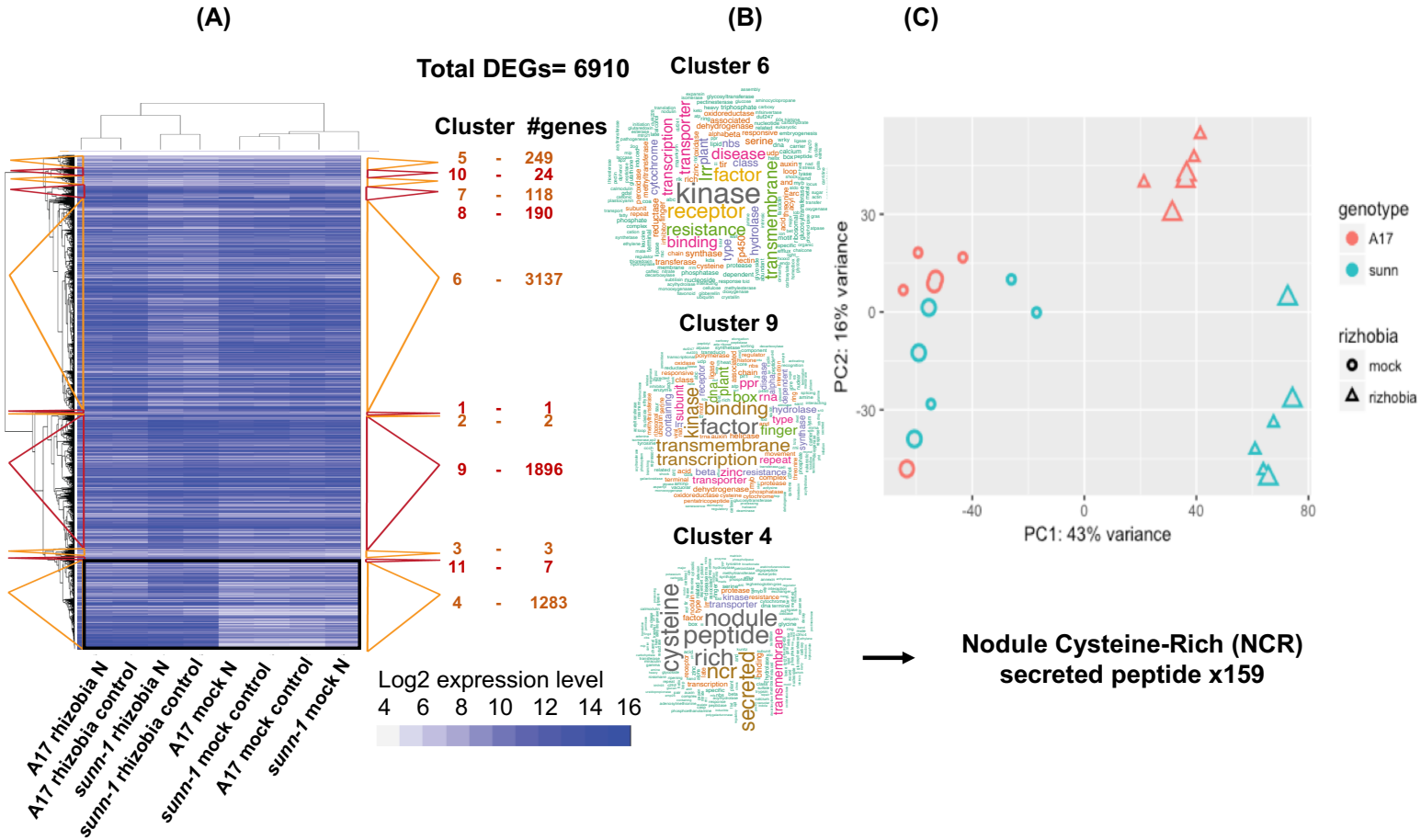


Figure 3.1.2: Heatmap showing hierarchical clustering showed division of 11 clusters with red/orange triangles representing the gene members in the cluster. Colour bar indicates log2 gene expression level ranging from 4-16. Gene expression pattern showed A17 and *sunn-1* microarray with three major clusters 4, 6 and 9. **(A)** Heatmap showing hierarchical clustering of the 6910 DEGs with their corresponding cluster numbers and genes. **(B)** Wordclouds depicting the most significant terms for the three largest clusters that represent 91.4% of the total DEGs. Gene enriched were: cluster 4- related to rhizobium responses Nodule cysteine rich peptides; cluster 6- related to N responses, kinase, transport, redox, calcium, nitrogen and UDP. See Appendix table 1 (C) Principal component analysis revealed distinct separation of mock and rhizobia response in A17 and *sunn-1* genotype.

Within cluster 6, an ortholog (*Medtr7g092930*) of the squamosa promoter-binding-like protein SPL9, involved in LR development of *A. thaliana* (Yu et al., 2015) was found to be co-expressed. This cluster also contains 10 Lateral Organ Boundaries (LOB-domain) genes, that typically have essential roles in integrating development in response to environmental changes (Xu et al., 2016). Wall associated kinase (11 genes) and expansin protein (10 genes) that mediate nodule cell walls (Giordano and Hirsch, 2004) were found. Out of three clusters, cluster 6 had higher number of auxin induced/responsive genes (37 genes) than in cluster 4 (6 genes) and 9 (20 genes). Aside from abundance of N-related genes found, cluster 6 also consisted of genes related with ethylene formation and response (16 genes), abscisic acid (5 genes), Myb-transcription factor (23 genes), WRKY family transcription factor (20 genes).

Genes in cluster 9 (1896 DEG), displayed distinctive stronger expression in *sunn-1* than in A17 independent of the N-treatment (Figure 3.1.2). A Low level of gene expression was found in A17 during N repletion, whether rhizobia was present or absent. This cluster consisted regulatory genes namely, kinase (95 genes), transmembrane (86 genes) and transcription (91 genes). Upregulation of the 35 LRR-kinases and 11 LysMs in this cluster could be related to the altered perception of rhizobia in the *sunn-1* mutant.

3.1.3 Fold change analysis highlighted a different magnitude of gene expression responses in A17 and *sunn-1*

To identify the strongest gene responses affected by varying levels of N and rhizobium inoculation, the fold change of the DEGs were analysed after applying a cut-off +1.5 and -1.5 and p-value 0.05. All the DEGs were assigned to 12 groups of samples that represents of (i) deplete and replete N (ii) rhizobium presence or absence samples in A17 and *sunn-1* (Table 3.1.3). Patterns of comparisons between treatments were generated separately for the defined groups to identify the strongest responsive genes as (i) Rhizobium effect, (ii) N effect and (iii) genotype effect. All the treatment sample groups and abbreviations used for classifying the groups (Table 3.1.3) for FC comparison are given in (Table 3.1.3).

For **rhizobium effect** (Table 3.1.3) in A17 deplete N (i.e. N deprived samples treated with 0.1 mM N) FC was analysed between mock treated AMLL and ARLL in A17, and SMLL and SRLL in mock treated *sunn-1*. Similarly, rhizobium effect at replete N (N deprived samples treated with 5 mM N) was calculated with FC between AMLH and ARLH in A17, and SMLH and SRLH in *sunn-1*.
For the **N effect** (Table 3.1.3) in absence or presence of rhizobia, FC comparison between mock treated plants of A17 at deplete and replete N was calculated (AMLL and AMLH), rhizobia-treated plants A17 as (ARLL and ARLH). Similarly, in *sunn-1* mutant FC comparison for N effect was calculated as (mock, Rhizobium absence- SMLL and SRLH) and (SRLL vs SRLH).
For the **genotype effect** (Table 3.1.3) FC patterns were calculated between AMLL and SMLL (*sunn-1* mutant effect in depleted N, -Rhiz), ARLL and SRLL (*sunn-1* mutant effect in depleted N, +Rhiz), AMLH and SMLH (*sunn-1* mutant effect in replete N, -Rhiz), between ARLH and SRLH (*sunn-1* mutant effect in replete N, +Rhiz) (data not shown).

Overall, rhizobium had stronger effect than N effect for both the genotype with greater number of differentially expressed genes of FC >1.5 or FC<-1.5 (Figure 3.1.3 A). Out of a total of 1099 strong-responding DEGs in A17, 96% (1066) were in rhizobia-inoculated plants. 502 genes in A17 showed differential expression when rhizobium was inoculated in an N deplete background (487 upregulated, 15 downregulated) and 558 genes in A17 were differentially expressed in an N replete background (552 genes upregulated, 6 genes downregulated) (Figure 3.1.3 B). The hypernodulant *sunn-1* mutant also showed a similar response with a prominent rhizobium effect during N depletion and repletion (Figure 3.1.3 A, B). Around 95% of DEGs (2280 genes out of 2396 DEGs) displayed a strong effect of rhizobium. However, the combinatorial effect of N and rhizobium in both the genotype was always stronger during N repletion than in N depletion (Figure 3.1.3 B).

Table 3.1.3: Group of samples that were compared together to calculate fold changes (Also in figure 3.1.3-B representing groups from 1-8). FC of the expression values of the differentially expressed genes between different samples (as indicated in the table) was calculated to identify genes responding to the effects listed in the table.

Response	**FC Calculated between samples**	
	A17	***sunn-1***
Replete N effect in -rhizobia	Group 1 AMLL vs AMLH	Group 2 SMLL vs SMLH
Replete N effect in +rhizobia	Group 3 ARLL vs ARLH	Group 4 SRLL vs SRLH
Effect of rhizobia at deplete N	Group 5 AMLL vs ARLL	Group 6 SMLL vs SRLL
Effect of rhizobia at replete N	Group 7 AMLH vs ARLH	Group 8 SMLH vs SRLH
SUNN effect at deplete N and -rhizobia	Group 9 AMLL vs SMLL	
SUNN effect at deplete N and +rhizobia	Group 10 ARLL vs SRLL	
SUNN effect at replete N and -rhizobia	Group 11 AMLH vs SMLH	
SUNN effect at replete N and +rhizobia	Group 12 ARLH vs SRLH	

When the N effect is considered, A17 and *sunn-1* showed a higher number of downregulated genes during N repletion than in N depletion in the absence of rhizobium (23 downregulated genes in A17, 91 downregulated genes in *sunn-1*) (Figure 3.1.3 B). N repletion in the absence of rhizobium had the highest effect in *sunn-1* with differential expression of 97 genes (91 downregulated, 6 upregulated) (Figure 3.1.3 B).

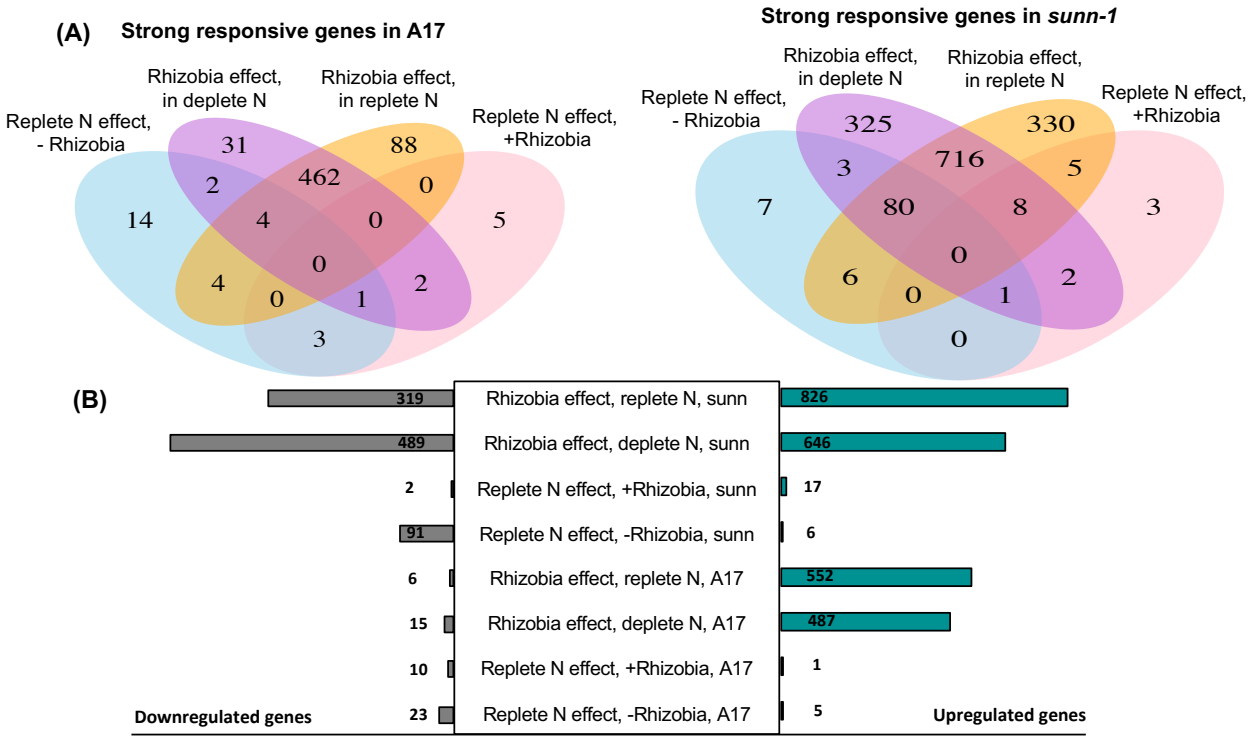


Fig 3.1.3: (A) Four set Venn diagram showing unique and shared genes that are strongly responsive to N and rhizobia in A17 and *sunn-1*. Sample groups represent highly regulated genes with FC >1.5 and FC <-1.5 (p-value < 0.05) in A17 and *sunn-1* for- replete N responsive genes in the absence of rhizobium; replete N responsive genes in the presence of rhizobium; rhizobia responsive genes in deplete N; rhizobia responsive genes in replete N. Also see table 3.1.3-1 . **(B)** Total number of DEGs that show significant upregulation or downregulation in A17 or *sunn-1*. The strongest responsive genes in each of the groups were selected based on FC >1.5 and FC <-1.5 to identify rhizobia effects, N effects and genotype effects in A17 and *sunn-1* from the sample group presented in table 3.1.3-1. FCs were calculated as (1) AMLH-AMLL; Group 1 (Replete N effect, -rhizobia in A17) (2) SMLH-SMLL; Group 2 (Replete N effect, -rhizobia in *sunn*) (3) ARLL-ARLH; Group 3 (Replete N effect, +rhizobia in A17) (4) SRLL-SRLH; Group 4 (Replete N effect, +rhizobia in *sunn*) (5) ARLL-AMLL; Group 5 (Rhizobia effect, deplete N in A17) (6) SRLL-SMLL; Group 6 (Rhizobia effect, deplete N in *sunn*) (7) ARLH-AMLH; Group 7 (Rhizobia effect, replete N in A17); (8) SRLH-SMLH; Group 8(Rhizobia effect, replete N in *sunn*). Genotype effect data for groups 9-12 not shown.

3.2 Discussion

This chapter mainly aimed to investigate the effect of nitrogen and rhizobium inoculation on root development and nodulation. For this work we used the autoregulation of nodulation mutant *sunn-1* because autoregulation mutants are disrupted in long-distance signalling AON and show nodulation independent phenotypes- increased LR density and hyper-nodulation (Wopereis et al., 2000). The hyper-nodulation mutant *sunn-1* was confirmed to still able to form nodules at a concentration of 5 mM NH_4NO_3 (N repletion), conditions where wild-type A17 nodulation was inhibited. In this work using root architecture studies in A17 and *sunn-1* during nodulation it was found that the two genotypes show different N-status-dependent phenotypic responses when seedlings were inoculated with rhizobium.

In general, primary root length was significantly shorter when rhizobia was present for both A17 and *sunn-1* and is independent of N availability. A study in *Arabidopsis* reported N-insensitive primary root elongation (Zhang et al., 1999). The reduction in the elongation of PR length as an effect of rhizobia is similar to a recent findings whereby rhizobium sp. IRBG74 was able to not only colonise *Arabidopsis* roots but also led to significant reduction in primary root growth (Zhao et al., 2017). Although the PR was shorter in both N conditions, the PR was even shorter in rhizobia-and-replete N conditions, highlighted a complex regulation and balance between nodule formation and RSA with an additive effect of N and rhizobium inoculation.

There was no significant difference in the lateral root (LR) number observed in our study in N-deplete conditions without rhizobia in A17 and *sunn-1*. This is consistent with findings from (Schnabel et al., 2005). However, *sunn-1* exhibited different characteristics in this work, with significantly higher LR numbers than A17 in N-replete conditions when either were mock- or rhizobia-inoculated. As well as there being more lateral roots, LR length was shorter in *sunn-1* in N-deplete rhizobium-inoculated conditions, showing evidence for the interaction of LR and nodule development. Underlying this effect is likely to be altered allocation of the photosynthates or C resources that inhibit LR outgrowth with increased nodule formation in *sunn-1*.

Root responses to N limited conditions are typically mediated by changes in PR, LR and root hair development; in total, root system architecture. To enable this, energy/carbon photosynthate is required to reduce the available N, assimilate nitrate and to transport nitrogenous compounds. Increased nodule formation in the *sunn-1* mutant does not translate into more N- fixation, as compared to WT (Mortier et al., 2012b), suggesting that the nodules are not very efficient. As N-fixation is a high energy costing process, C photosynthate supply for inefficient nodules that do not fix N can be cut-off (Oono et al., 2011). This could be one of the reasons underlying the RSA response in the *sunn-1* mutant - shorter LR in the presence of rhizobium possibly could be a resource balancing mechanism to counteract hypernodulation.

N uptake in *Medicago truncatula* had been reported in the form of three main pathways NH_4, NO_3 and NO_2 (Ruffel et al., 2008). However, the common regulation of such nutrient uptake is repressed by local and systemic signalling that depends on downstream N metabolite accumulation. Based on evidence of differential N acquisition pathway response, the transcriptomic investigation here could suggest that systemic signalling acts upon gene regulatory networks in response to external N supply from the environment (Jeudy et al., 2010).

In addition to the phenotypic response presented above, microarray expression profiling was used to understand the effect of rhizobium and N in *M. truncatula*. Among rhizobia responsive DEGs, there was a strong upregulation in response to the combinatorial effect of N repletion and rhizobia (552 genes out of 558) in A17 (826 genes out of 1145) in *sunn-1*.

Hierarchical clustering from the transcriptomic analysis suggested the presence of three major distinctive clusters (4, 6 and 9). Cluster 4 represented genes involved mainly in the nodulation pathway that responded to rhizobia effect. Cluster 6 comprised of N-related genes that had an N effect, while cluster 9 represented a genotype effect related to *sunn-1*. Out of these three clusters, cluster 4 and 6 were of particular interest as this thesis aimed to study the combinatorial effect and regulation of N and rhizobium in Medicago.

As highlighted previously (section 3.1.2), a greater number of genes was found to be induced in A17 than in *sunn-1* in cluster 6 during N-repletion. This cluster was found to be comprised mainly of N-related genes. RNAseq

transcriptome analysis of the nitrate impact on root nodules in A17 during 4-28hr of NO_3 exposure found expression of two genes (Medtr7g069640, Medtr8g095040) belonging to ammonium transporter family proteins (Cabeza et al., 2014). We also find these two genes to be expressed at 6 hour N-repletion in the presence of rhizobium in A17. However, significant downregulation with FC (FC<-1.5) of these genes was seen in *sunn-1* from our expression data.

The expression pattern of the Medtr3g035860 gene, encoding for glutamine synthetase (GS), the first enzyme in N assimilatory pathway (Seabra et al., 2013) in this cluster is crucial to note. Whilst there was high expression of GS gene *Medtr3g035860* in wild-type A17 in all conditions, there was a significant downregulation of this gene expression in *sunn-1* (FC<-1.5) in N-depletion in the presence of rhizobium. These findings suggested that nitrogen assimilation could be more efficient in A17 than in *sunn-1* in the presence of rhizobia, potentially leading to inhibition of nodulation in A17 in N-replete conditions. Significant downregulation of the *GS* gene in *sunn-1* also supports insensitivity of the hypernodulating mutant to external N status (5 mM) that is typically high enough to repress nodulation.

The abundance of phytohormonal transcripts, especially auxin (37 genes) and ethylene responsive (16) genes in cluster 6 also implicate auxin and ethylene as key players to regulate RSA in the control of nodulation. Auxin acts as mediator in the AON pathway within the mechanism involving a shoot-derived long distance signal to regulate nodule and LR development (Jin et al., 2012). Expression of auxin related genes in cluster 6 (upregulated in A17 but not in *sunn-1*) is consistent with the phenotypic differences in A17 and *sunn-1* in N replete rhizobia-inoculated conditions (shorter PR phenotype in A17 when compared to *sunn-1* (Figure 3.1.1). Longer LRs in A17 when rhizobia-inoculated could be explained by the induction of LOB-domain genes in this condition (Figure 4C). These genes could be key in the regulation of LR length, integrating the internal and external N signals to result in an appropriate or optimal developmental response.

Out of 37 auxin related genes, 13 genes were found to be significantly downregulated with FC <-1.5 in the *sunn-1* mutant on N depletion and in the

presence of rhizobium. N repletion in the presence of rhizobium also led to downregulation of 3 auxin related genes (*Medtr4g129510*, *Medtr7g079720*, *Medtr4g010340*, FC<-1.5). LR development mediated by auxins generally involve interference from ethylene by altering auxin transport and we found evidence of this in our data (Ivanchenko et al., 2008; Lewis et al., 2011). Studies from (Stepanova et al., 2007; Swarup et al., 2007) indicate that ethylene and its precursor could inhibit root elongation synergistically. Significant downregulation of such ethylene associated genes (*Medtr2g015040*, *Medtr1g101550*, *Medtr2g025120*) was found in cluster 6 (FC < -1.5) in response to N-depletion and rhizobium in the *sunn-1* mutant. Such cross-talk of ethylene and auxin effects could be a possible explanation for the reduced primary root growth seen in *sunn-1* in presence of rhizobia. Downregulation of expansin and extension family genes in *sunn-1* in the presence of rhizobium could also be likely associated with PR growth inhibition in this condition. This is in line with a study in soybean where rapid PR elongation during root development was found to be correlated with maximum level of an expansin gene, *GmEXP1* (Lee et al., 2003).

Of the major three clusters, cluster 4 exhibited a prominent rhizobia response with many genes involving in the nodulation pathway (as described in section 3.1.2). Aside from nodulins and leghaemoglobins expressed in this cluster, cluster 4 is of particular interest as it includes many members of the massive family of nodule-specific genes encoding cysteine rich peptides. Out of 185 nodule-specific cysteine rich (NCRs) peptides genes expressed on our microarray data, 165 NCR genes were found to be differentially expressed (DE). Legumes in the inverted repeat lacking clade (IRLC) of legumes (including *Medicago truncatula*) contain more than 500 members of NCR multi-gene family. Of the DE NCR genes, 159 were in cluster 4. Only three NCRs (Medtr1g042850.1, Medtr3g062880.1, Medtr5g057460.1) belonged to 9; cluster 7 had two NCRs (Medtr3g084820.1, Medtr6g006240.1) and there was only NCR in cluster 6 (Medtr5g057910.1).

Cluster 4 thus showed a unique feature whereby the 159 NCR genes and other primary genes involved in nodulation pathway including as nodulins,

leghaeomglobins and lys-motif were significantly upregulated (FC>1.5, p<0.05) in A17 and *sunn-1* in response to rhizobium and independent of N status.

Most previous transcriptomic and proteomic studies aimed at understanding the function of NCRs have been confined to investigating their antimicrobial activity controlling terminal differentiation of the invading bacteria. However, given the extensive sequence diversity of the NCR family over different expression stages, the answer to why IRLC legumes including *M.truncatula* employ a large gene family for this function remains incomplete. One hypothetical model suggested that NCRs keep the over-proliferation of the bacteroids in check to enable more efficient legume-rhizobium symbiosis (Pan and Wang, 2017). The functional investigation on three NCRs of focus, NCR 247, NCR 169 and NCR 211, out of 600 NCR genes represents only a minute portion of this large size gene family (Farkas et al., 2014; Horvath et al., 2015; Kim et al., 2015). Due to their occurrence in large numbers from genome amplification, it is possible that these NCR genes have evolved gain of novel functions beyond just bacterial regulation and our findings that they are rhizobium-regulated in A17 and *sunn*, invariant of N status is intriguing and worthy of further study.

Chapter 4: Promoter analysis reveals novel motifs in the NCR gene family and putative regulatory transcription factors

4.1 Introduction

Genome wide expression profiling with microarray was used to study the pattern of gene expression changes in A17 and *sunn-1*. Using a fold change cut-off 11 clusters of DEGs were identified in 11 groups, enabling delineation of the strongly regulated genes under N or rhizobium or genotype effects (previously listed on table 3.1.3, also see Figure 3.1.3 A, B). These genes were hypothesised to be part of the regulatory control mechanism that governs gene expression during nodulation. This regulation could be at the level of combinatorial effects of N and rhizobium that involves complex cross-talk between external N-status availability and rhizobium recognition and interaction.

Gene expression regulatory mechanisms are complex, involving frequent DNA-protein interactions that control the level of mRNA expression. One of the most important initial steps in gene expression, transcription, is a fundamental process under regulation by many key components involving interactions between the gene promoter, cis-acting regulatory binding sites, transcription factor(s) and other distal elements including enhancers or silencers.

To gain a better understanding of transcriptional regulation that controls activation or suppression of the genes regulated by rhizobium and nitrogen in our experimental conditions, we sought to study the promoter region of the DEGs in the 12 regulatory groups (listed on table 3.1.3-1, also see Figure 3.1.3 A, B) identified from our microarray analyses. With TF being highly specific in binding to its cis-acting DNA motif, identification of motifs on the upstream region of the differentially expressed and co-regulatory group of genes becomes informative in identifying upstream transcriptional regulators.

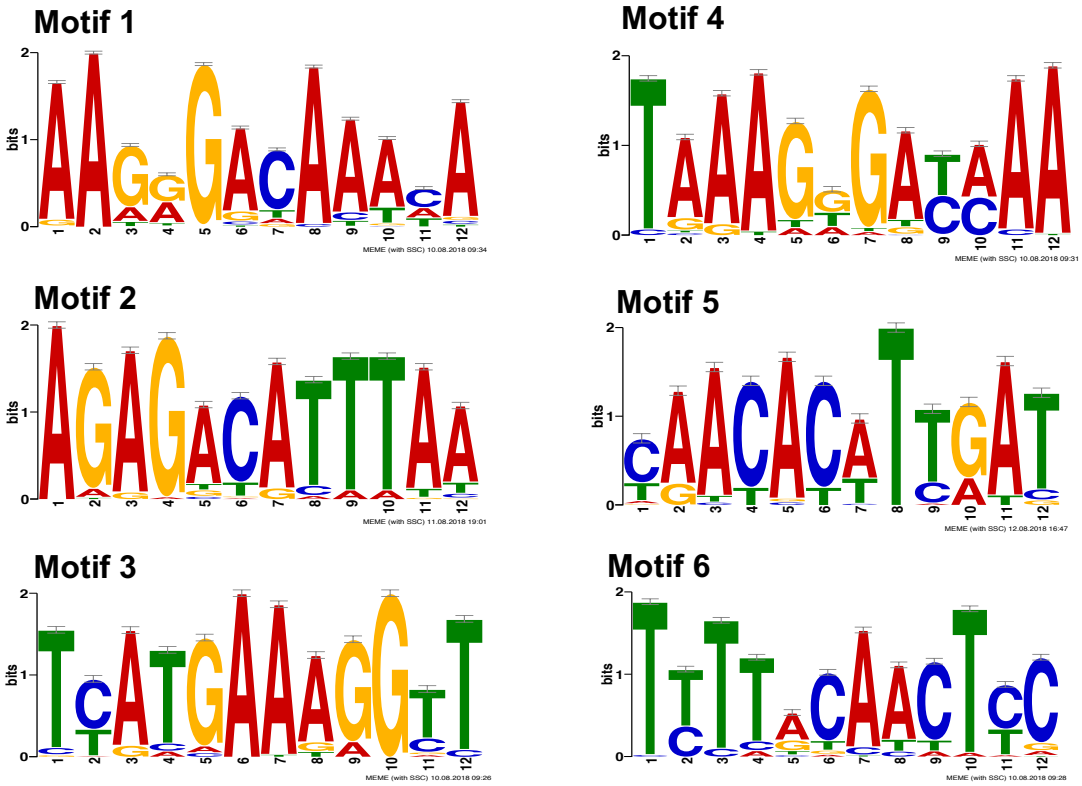
4.1.1 Novel motifs are present on the upstream region (promoter) of NCR genes

MEME suite version 5.0.1 (Bailey et al., 2015) was used to investigate the upstream regions of the DEGs that were regulated by rhizobium and nitrogen responses in A17 and *sunn-1*. The run was performed on the 12 groups of DEGs and significant motifs were selected based on a criteria that motifs were found in >20% of promoters with information content as bits size (range from 0-2) indicating conservation of motifs (Figure 4.1.1), positional bias (p-value) and an E-value < 0.001, indicating motif was unlikely due to random artifact or chance. A previous motif search (Roxanna Bonyadi, Thesis) was performed on the Mt3.5 genome in 200 bp, 500 bp and 1000 bp upstream regions of the genome using MEME-LaB tools (Brown et al., 2013). It showed that motifs were clustered within the 500 bp region. Hence, the motif analyses in this thesis were focused on the upstream 500 bp, and using the Mt4.0 genome as reported by (Nallu et al., 2013).

Based on the above criteria, MEME-suite analyses on the upstream region of differentially expressed promoter genes identified six significant novel motifs, (TCATGAAAGGTT, TATAAAGTGATCA, CAACACATTGAT, AGAGACATTTAA, TTTTACAACTCC (Figure 4.1.1). The promoters where these motifs were found included of 66-81% NCR genes for all but the motif AAGGGACAAACA (found in 46% of NCRs) (Figure 4.1.1 and table 4.1.1). These motifs were identified in the group of differentially expressed genes that were rhizobia-regulated at low N in A17 (group 5: motifs 1, 2 and 3) and *sunn-1* (group 6: motifs 1, 2 and 3); and rhizobia-regulated at high N in A17 (group 7: motifs 1, 2, 3, 4, 5 and 6) and *sunn-1* (group 8: motif 1) as given in (Figure 4.1.1 and table 4.1.1). For the group division see chapter 3, section 3.1.3, table 3.1.3.

Aside from NCRs and proteins with unknown function, other gene promoters with these motifs were mainly late nodulins, peptide/nitrate transporter proteins, leghemoglobins, nodule inception protein and *ENOD18*. Hence, the motifs were found to be present in promoters of around 90% of the genes that are involved in the nodulation pathway. Significant motifs with E-

values (Table 4.1.1-1) and their logos as discovered (Figure 4.1.1) within the groups of co-expressed genes are given. The presence of motifs from the *de novo* motif discovery in the promoter upstream region suggests putative cis-regulatory sites in the NCR promoters for TF binding and thus regulation.



Motifs found in rhizobia effect for A17 & *sunn-1* at deplete and replete N

Motifs found in rhizobia effect for A17 & *sunn-1* at high N

Fig 4.1.1 De novo motif discovery in promoters of N/rhizobium-regulated genes. Logos of the significant motifs discovered in the DEGs and co-expressed genes of the sample group from MEME suite analysis. Bits in motif logos correspond to the conserved frequencies of the nucleotides A/T/G/C. Three motifs were found to be overrepresented in the sample group 5, 6 and 7 with rhizobia effect at both deplete and replete N condition. Motifs have been assigned number as 1-3. Table 4.1.1

Table 4.1.1 Percentage of differentially expressed NCRs that contain six motifs. Motif sequence and E-values (number of estimate that motif occurs by chance) have been shown along with the % of NCR promoters that contain the motifs. NCRs with motif 1, 2 and 3 were found to be expressed under rhizobia effect in A17 and *sunn-1*, deplete or replete N. NCRs with motif 4, 5 and 6 were found to be expressed in rhizobia effect in A17 and *sunn-1* at replete N condition.

Motifs	NCR promoters out of all the promoters	% of NCRs	Conditions
Motif 1 AAGGGGACAAACA E= 1.6e-129	111/240	46	Rhizobia effect (A17 and *sunn-1*), deplete/replete N condition
Motif 2 AGAGACATTTAA E= 6.8e-080	60/77	79	
Motif 3 TCATGAAAGGTT E= 1.5e-050	159/214	74	
Motif 4 TAAAGGGATAAA E=8.0e-076	57/70	81	Rhizobia effect (A17 and *sunn-1*), replete N condition
Motif 5 CAACACATTGAT E=1.4e-020	45/62	73	
Motif 6 TTTTACAACTCC E=1.4e-054	42/64	66	

4.1.2 Motif enrichment testing

As identification of motif discovery is affected by the choice of gene sequences or nucleotide composition in background set, we sought to test the presence of motifs in the promoters in the remainder of the genome. To determine the abundance of these motif consensus predictions, position weight

matrix scanning for individual motif occurrences were performed using FIMO (Find Individual motif occurrences) (Grant et al., 2011). This method allowed us to establish the level of motif enrichment by searching the enumerated motifs generated from MEME against the set of promoter genes drawn from non-NCR genes in transcriptomic data and non-expressed genes in Medicago genome.

Using every occurring k-mer (12 in our case) as our motif length, FIMO was used to determine the matched position of each motif based on log-likelihood scores. We hypothesised that NCR gene promoters were enriched with the novel motifs identified that could serve as putative binding sites for TF(s) as many of the co-regulated genes with motifs were found to be NCRs. The gene expression threshold (log2 hybridisation signal) for each of the treatment conditions in A17 and *sunn-1* microarray transcriptomic data was determined to be in the range of 6-8 (data not shown). All the Mt4.0 gene transcript that matched probe ID with log2 hybridisation signal greater than their corresponding threshold values were considered expressed. Hence, all the Mt4.0 185 NCR genes found on microarray data were considered for the motif enrichment scan.

To determine if the motif enrichment scores in expressed NCR promoter regions were similar in other genes of the Medicago genome, we performed FIMO on three datasets of randomised genes, compared to the 185 regulated NCR genes. The background datasets were; (i) NCRs absent on microarray (ii) expressed non-NCRs (iii) non-expressed and non-NCRs. All the six motifs were found to be enriched in the expressed NCR promoters compared to all background datasets (Figure 4.1.2). Motif occurrence sites in the negative backgrounds (expressed non-NCRs and non-expressed non-NCRs) were low ($p < 1E^{-4}$). The mean motif occurrences score for the enrichment can be summarised as; expressed NCRs > NCRs absent on microarray > expressed non-NCRs > non-DE and non-NCRs. This was the case for all six motifs, suggesting that the motifs are more prevalent in *NCR* genes.

FIMO scan comparison with the background of NCRs absent on microarray (Figure 4.1.2), motif 1, 3, 4 and 6 were found to be highly enriched in the expressed 185 NCRs. We also found that motif 2 and 5 were present in non-expressed NCR genes (Figure 4.1.2). However, the highest enrichment

with exact match was found to be in the expressed *NCR* genes at the threshold p-value < 1E-4. The negative background sets of expressed non- NCRs/non-expressed and non-*NCR* genes were found to contain a few motif sites, but these were not enriched in comparison to *NCR* promoter gene background (Fig 4.1.2).

The six motifs found to be present in the *NCR* promoter genes are given on Table 4.1.2-1. Promoters of 146-149 *NCR* genes contained motif 1 and 4 closely followed by 133 NCR promoters containing motif 2. Motif 5 and 6 were found in 94-95 number of *NCR* genes Table 4.1.2-1. The individual motif occurrence present in all the 185 expressed NCR genes are given in Table 4.1.2-2. Out of all the 185 expressed NCRs, 40 NCRs were found to contain all the six motifs (Table 4.1.3-1).

Table 4.1.2-1: Frequency of motif occurrence found in all the promoters of 185 expressed NCRs

Motifs	**No motifs present**	**Motif occurrence =1**	**Motif occurrence =2**	**Motif occurrence = 3 or more**	**Total no. of promoters with the motifs**
Motif 1	39	57	55	34	146/185
Motif 2	52	92	31	10	133/185
Motif 3	67	70	37	11	118/185
Motif 4	36	63	46	40	149/185
Motif 5	90	70	23	2	95/185
Motif 6	91	80	10	4	94/185

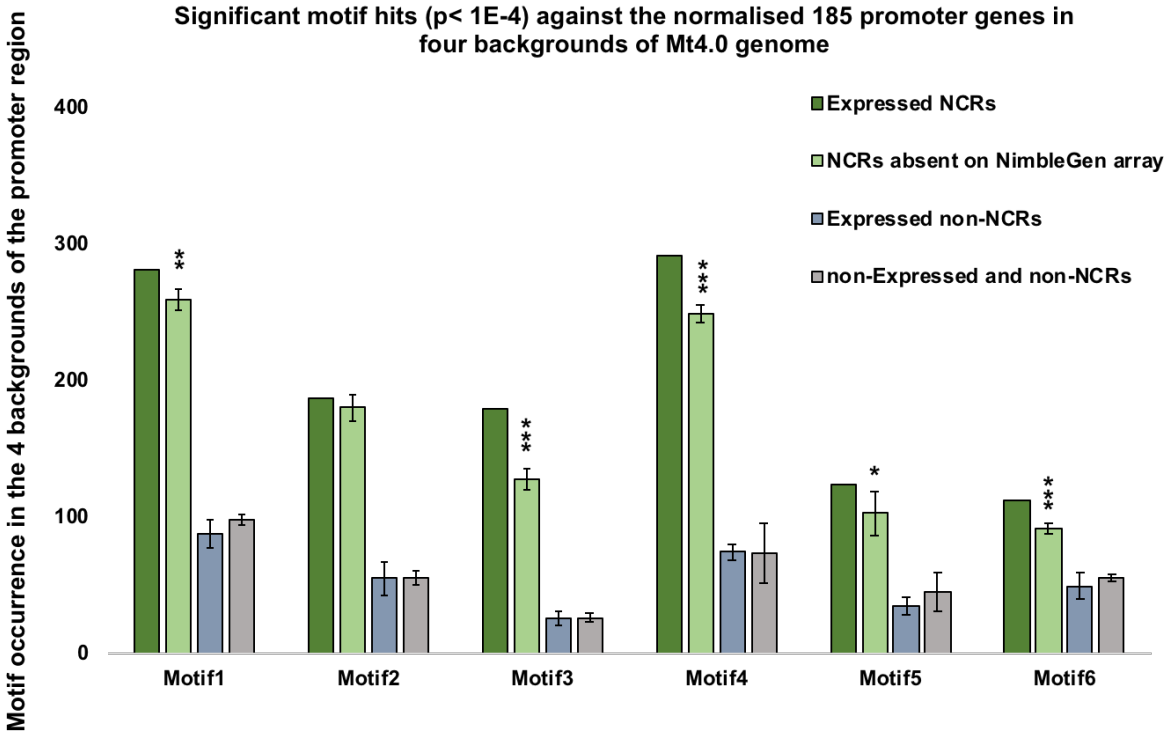


Figure 4.1.2: Motif enrichment scan in four backgrounds of Mt 4.0 genome to ask if motifs are a feature of regulated *NCR* gene promoters. Four backgrounds are expressed NCRs, NCRs absent on microarray, expressed non-NCRs, non-expressed and non-NCRs. The promoters within each group (except those in the ' expressed NCRs' list were randomly selected from the genes with those features in the remainder of the Mt4.0 genome. Each group was chosen to be a total of 185 to be equal to the number of NCR genes expressed on the array. Error bars represent standard deviations for three randomised runs (n=3)

Table 4.1.2-2: Individual motif occurrences in all the 185 expressed NCR genes from the microarray transcriptomic data. Number represents the frequency of the motifs occurring in the NCR upstream promoter region [-1,-500]. Different shades of colour and with bars corresponds to its frequency. All the Medicago gene IDs were arranged in the order of motif 2 occurrence from frequency largest to smallest.

(A)

Expressed NCRs	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6
Medtr3g021040.1	2	4	1	2	2	2
Medtr3g069830.1	2	4	0	2	1	1
Medtr7g016440.1	2	4	1	4	1	0
Medtr1g042200.1	1	3	0	0	1	0
Medtr1g058880.1	0	3	0	2	1	0
Medtr2g104570.1	1	3	0	0	1	0
Medtr3g015940.1	2	3	3	2	1	0
Medtr3g033100.1	2	3	0	2	2	1
Medtr5g063580.1	3	3	2	2	1	0
Medtr5g066750.1	0	3	1	1	0	1
Medtr1g042910.1	1	2	3	0	0	0
Medtr1g046020.1	2	2	0	1	2	3
Medtr1g074860.1	3	2	0	3	1	1
Medtr3g010630.1	3	2	1	3	2	0
Medtr3g014720.1	3	2	1	2	1	0
Medtr3g015870.1	3	2	2	3	1	0
Medtr3g020930.1	2	2	1	1	0	1
Medtr3g020980.1	1	2	2	2	1	1
Medtr3g021050.1	1	2	0	1	1	1
Medtr3g033260.1	3	2	0	1	2	1
Medtr3g033930.1	2	2	0	2	0	1
Medtr3g069870.1	2	2	0	2	1	1
Medtr3g071360.1	2	2	0	1	0	0
Medtr4g027000.1	2	2	0	1	1	0
Medtr4g031380.1	0	2	1	1	1	1
Medtr4g031430.1	1	2	0	2	1	1
Medtr4g031520.1	0	2	2	3	1	2
Medtr4g060730.1	4	2	0	3	1	1
Medtr5g048310.1	2	2	1	0	1	1

(B) Table 4.1.2-2 (continued)

Expressed NCRs	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6
Medtr5g056710.1	2	2	1	1	0	0
Medtr5g057460.1	1	2	0	3	0	0
Medtr5g059740.1	3	2	0	4	1	0
Medtr5g061160.1	1	2	1	1	0	1
Medtr5g061400.1	0	2	2	3	0	1
Medtr5g063600.1	3	2	0	2	1	0
Medtr5g069100.1	2	2	0	2	1	1
Medtr5g070410.1	2	2	1	3	0	2
Medtr5g072070.1	3	2	1	2	0	1
Medtr7g045910.1	2	2	0	2	1	1
Medtr7g071220.1	1	2	1	1	2	0
Medtr7g071310.1	4	2	1	2	0	0
Medtr1g042850.1	1	1	0	1	0	1
Medtr1g042940.1	0	1	0	0	1	0
Medtr1g044500.1	4	1	0	1	2	1
Medtr1g075500.1	1	1	0	2	0	0
Medtr1g092720.1	2	1	0	0	0	1
Medtr2g045290.1	2	1	1	3	0	1
Medtr2g050060.1	1	1	0	2	1	1
Medtr2g060960.1	0	1	0	0	1	0
Medtr2g063470.1	2	1	2	0	2	2
Medtr2g083280.1	4	1	2	2	4	0
Medtr3g016020.1	3	1	0	4	0	0
Medtr3g016090.1	2	1	2	3	2	0
Medtr3g016130.1	2	1	1	4	1	0
Medtr3g020920.1	1	1	1	1	1	1
Medtr3g020950.1	3	1	1	2	2	1
Medtr3g028380.1	2	1	2	2	2	1
Medtr3g028440.1	1	1	1	1	3	1
Medtr3g030060.1	1	1	1	1	2	0
Medtr3g031340.1	1	1	2	0	1	1
Medtr3g034220.1	1	1	1	0	0	0
Medtr3g061750.1	2	1	1	1	1	1
Medtr3g062810.1	1	1	0	1	0	1
Medtr3g062820.1	2	1	1	1	0	1
Medtr3g062830.1	3	1	0	3	1	0
Medtr4g014790.1	1	1	0	4	0	0
Medtr4g026680.1	3	1	0	3	1	1
Medtr4g026750.1	2	1	2	2	0	0
Medtr4g031900.1	1	1	1	1	1	0

(C) Table 4.1.2-2 (continued)

Expressed NCRs	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6
Medtr4g031900.1	1	1	1	1	1	0
Medtr4g033830.2	2	1	2	2	0	0
Medtr4g053180.1	0	1	1	1	1	3
Medtr4g057120.1	0	1	1	0	1	1
Medtr4g057160.1	1	1	0	2	1	1
Medtr4g060590.1	1	1	1	2	0	2
Medtr4g060610.1	1	1	1	2	0	1
Medtr4g060650.1	0	1	3	1	0	0
Medtr4g060660.1	2	1	1	3	0	0
Medtr4g060720.1	0	1	2	2	1	1
Medtr4g070690.1	0	1	1	0	1	1
Medtr4g071890.1	1	1	3	0	0	0
Medtr5g026070.1	1	1	0	1	0	1
Medtr5g026080.1	3	1	2	2	1	1
Medtr5g037780.1	2	1	3	1	1	1
Medtr5g040380.1	0	1	1	0	0	0
Medtr5g047670.1	2	1	0	3	0	0
Medtr5g056360.1	2	1	0	1	1	0
Medtr5g056760.1	1	1	2	1	0	1
Medtr5g056890.1	1	1	0	1	0	1
Medtr5g057910.1	3	1	0	4	0	0
Medtr5g059670.1	3	1	0	2	2	0
Medtr5g061060.1	2	1	0	2	1	0
Medtr5g061120.1	3	1	1	3	1	1
Medtr5g061800.1	1	1	1	1	0	1
Medtr5g062630.1	2	1	0	1	0	1
Medtr5g063460.1	2	1	0	3	1	0
Medtr5g063490.1	1	1	0	3	0	1
Medtr5g063520.1	0	1	0	0	1	1
Medtr5g063780.1	2	1	2	2	1	1
Medtr5g063890.1	2	1	1	1	0	0
Medtr5g064860.1	2	1	1	3	0	2
Medtr5g068810.1	1	1	2	3	0	2
Medtr5g069500.1	1	1	2	3	2	1
Medtr5g071880.1	0	1	0	2	0	0
Medtr5g072310.1	1	1	1	1	1	2
Medtr5g072420.1	2	1	1	3	0	0
Medtr5g072450.1	2	1	0	1	1	0
Medtr5g095620.1	0	1	1	0	0	0
Medtr6g006240.1	2	1	2	0	0	0

(D) Table 4.1.2-2 (continued)

Expressed NCRs	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6
Medtr6g006240.1	2	1	2	0	0	0
Medtr6g006250.1	3	1	2	0	0	1
Medtr6g006350.1	1	1	0	2	1	1
Medtr6g044730.1	1	1	2	2	0	0
Medtr6g055700.1	2	1	2	1	1	0
Medtr6g057520.1	0	1	1	3	1	1
Medtr6g060320.1	0	1	1	0	1	0
Medtr7g008940.1	6	1	1	2	0	0
Medtr7g008970.1	1	1	1	3	1	0
Medtr7g010200.1	1	1	1	1	0	1
Medtr7g027180.1	4	1	2	1	0	0
Medtr7g028550.1	1	1	1	2	1	1
Medtr7g029760.1	2	1	1	1	1	0
Medtr7g032720.1	2	1	1	0	0	0
Medtr7g032820.1	1	1	4	2	0	0
Medtr7g037690.1	1	1	2	2	2	1
Medtr7g045410.1	4	1	1	3	0	0
Medtr7g045520.1	0	1	2	1	1	1
Medtr7g051290.1	1	1	3	1	0	0
Medtr7g051320.1	3	1	2	1	0	1
Medtr7g052640.1	2	1	3	0	0	0
Medtr7g071690.1	1	1	1	0	2	0
Medtr7g071720.1	3	1	1	1	2	0
Medtr7g071720.2	3	1	2	0	2	0
Medtr7g071720.3	3	1	3	1	2	0
Medtr8g036850.1	3	1	2	1	2	0
Medtr2g008910.1	1	0	2	1	1	0
Medtr3g010490.1	1	0	2	4	0	1
Medtr3g014260.1	0	0	4	1	1	1
Medtr3g025420.1	1	0	0	0	0	0
Medtr3g027180.1	2	0	0	2	0	1
Medtr3g031320.1	1	0	0	1	0	2
Medtr3g033510.1	0	0	0	1	0	0
Medtr3g033700.1	0	0	1	0	1	0
Medtr3g052100.1	0	0	1	0	0	0
Medtr3g062880.1	2	0	0	1	1	1
Medtr3g065050.1	2	0	1	2	1	1
Medtr3g065690.1	0	0	1	1	0	1
Medtr3g065710.1	1	0	0	2	0	1
Medtr3g084820.1	2	0	1	2	1	0

(E) Table 4.1.2-2 (continued)

Expressed NCRs	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6
Medtr3g084820.1	2	0	1	2	1	0
Medtr3g084910.1	4	0	3	6	0	1
Medtr4g015750.1	0	0	1	1	0	1
Medtr4g033830.1	0	0	2	1	0	0
Medtr4g033900.1	1	0	1	1	1	0
Medtr4g057100.1	2	0	0	2	0	0
Medtr4g059900.1	2	0	0	3	0	1
Medtr4g060700.1	1	0	0	0	0	0
Medtr4g065390.1	1	0	0	1	0	0
Medtr4g100690.1	0	0	0	1	0	2
Medtr5g055370.1	5	0	1	8	0	1
Medtr5g058510.1	0	0	1	0	1	0
Medtr5g059440.1	0	0	2	0	2	1
Medtr5g061640.1	0	0	1	2	0	0
Medtr5g062510.1	0	0	1	1	0	1
Medtr5g069530.1	2	0	0	1	0	0
Medtr5g073530.1	1	0	1	3	0	3
Medtr5g073580.1	2	0	1	2	1	3
Medtr5g076040.1	1	0	1	3	0	1
Medtr5g095590.1	2	0	1	1	0	1
Medtr6g044700.1	2	0	2	3	2	0
Medtr6g055160.1	1	0	2	0	0	1
Medtr6g060370.1	0	0	2	0	1	1
Medtr6g061820.1	0	0	0	1	0	1
Medtr7g009040.1	0	0	2	1	1	0
Medtr7g016470.1	1	0	1	1	1	0
Medtr7g021640.1	2	0	1	0	0	0
Medtr7g025060.1	0	0	1	0	0	0
Medtr7g027120.1	3	0	0	3	1	1
Medtr7g037410.1	0	0	0	2	0	0
Medtr7g050990.1	1	0	0	1	1	1
Medtr7g070090.1	1	0	1	1	2	0
Medtr7g080850.1	1	0	2	0	0	1
Medtr7g084820.1	3	0	0	4	0	0
Medtr7g102160.1	2	0	0	1	0	1
Medtr7g102170.1	0	0	0	0	0	0
Medtr8g036830.1	0	0	1	0	0	0
Medtr8g069130.1	0	0	0	1	0	0
Medtr8g069310.1	1	0	1	3	0	0

Table 4.1.3-1: All the 40 expressed NCRs that contain all the six motifs. Numbers and shades indicate the frequency of motif occurrence in their promoter region

Expressed NCRs	Motif 3	Motif 4	Motif 1	Motif 5	Motif 2	Motif 6
Medtr3g069830.1	1	2	2	1	4	1
Medtr3g033100.1	1	2	2	2	3	1
Medtr5g073530.1	3	3	1	3	3	3
Medtr5g073580.1	2	2	2	1	3	3
Medtr3g031320.1	1	1	1	1	2	2
Medtr3g033260.1	1	1	3	2	2	1
Medtr3g069870.1	1	2	2	1	2	1
Medtr4g060730.1	3	3	4	1	2	1
Medtr5g061160.1	1	1	1	1	2	1
Medtr5g069100.1	2	2	2	1	2	1
Medtr5g070410.1	3	3	2	3	2	2
Medtr5g072070.1	3	2	3	2	2	1
Medtr7g045910.1	1	2	2	1	2	1
Medtr3g010490.1	2	4	1	4	1	1
Medtr3g020920.1	2	1	1	1	1	1
Medtr3g027180.1	2	2	2	2	1	1
Medtr3g061750.1	1	1	2	1	1	1
Medtr3g065710.1	1	2	1	2	1	1
Medtr3g084910.1	1	6	4	6	1	1
Medtr4g026680.1	1	3	4	1	1	1
Medtr4g057160.1	1	2	1	1	1	1
Medtr4g059900.1	3	3	2	3	1	1
Medtr4g060590.1	1	2	1	2	1	2
Medtr4g060610.1	2	2	1	2	1	1
Medtr4g060720.1	2	2	1	1	1	1
Medtr5g056890.1	2	1	1	1	1	1
Medtr5g062630.1	2	1	2	1	1	1
Medtr5g063490.1	1	3	1	3	1	1
Medtr5g063780.1	2	2	2	1	1	1
Medtr5g064860.1	1	3	2	3	1	2
Medtr5g068810.1	4	3	1	3	1	2
Medtr5g069500.1	1	3	1	2	1	1
Medtr5g072310.1	1	1	1	1	1	2
Medtr5g076040.1	2	3	1	3	1	1
Medtr7g010200.1	1	1	1	1	1	1
Medtr7g028550.1	1	2	1	1	1	1
Medtr7g037690.1	1	2	2	2	1	1
Medtr7g045520.1	1	1	1	1	1	1
Medtr7g050990.1	2	1	2	1	1	1
Medtr7g051320.1	2	1	3	1	1	1

4.1.3 Multiple sequence alignment reveals motif conservation

The MEME algorithm can be used to identify motifs based on an expectation-maximisation or an enumerative approach and can be used to generate a list of over-represented consensus sequences from a co-regulated set of gene promoters. As MEME outputs all the oligomers that occur more than the expected with a prior user-specified motif length, fixed motif widths based on rigid background nucleotide composition are often generated. Such user customised motif length may not necessarily be true representative of regulatory binding sites. Hence, to dissect the discovered motifs for homologies and conservation amongst promoters we sought to perform multiple sequence alignment to ask if the motifs were individual or part of longer stretches of conserved sequence.

NCR genes have been reported as one of gene families to have significant divergence from a whole-genome study of nucleotide study (Branca et al., 2011). Such extensive divergence among the NCR promoters made the creation of an accurate sequence alignment difficult, with low conservation across the 185 differentially expressed NCR promoters. And visualisation of all the NCR promoters generated scattered motif conservation patches. Mathematically, it is also very difficult to generate best global or local alignment unless gaps are introduced (Mount, 2008). Thus to produce a representative of an accurate alignment to build an effective Hidden Markov Model (HMM), eight DE-NCR genes were selected based on the occurrence of all the six motif sites and empirical observation. The upstream sequences [-1,-500] of these genes were aligned using the EBI tools MAFFT (Multiple alignment using fast fourier transform) (Katoh et al., 2002). The aligned sequences with open gap and gap extension penalty (as described in method section (2.9.3) was visualised for conservation in Jalview (Waterhouse et al., 2009) and Genious version 11.0.2 https://www.geneious.com/features/genome-alignment/.

The alignment visualisation in Genious (version 11.0.2) (Figure 4.1.3) revealed motifs conservation in the NCR upstream region. The alignment (Figure 4.1.3) showed that motifs 4, 1 and 5 were found to be overlapping, indicating that they could belong to a bigger motif region consisting of 39-40 nucleotides (see alignment figure 4.1.3). Notably, motif 3, 2 and 6 were

observed to occur separately with a TATA site (highlighted in yellow box) closely following motif 6 (Figure 4.1.3, highlighted in red boxes). We also found scattered stretch of conserved elements next to motif 2 (TAATGATTT) and motif 6 (CTATTTAATT) outside our six motif sites (Figure 4.1.3, highlighted in purple boxes). In summary, the alignment provided us an overview of the conserved sites of motif landscape in the NCR promoter region.

The combinatorial occurrence of the six motifs were determined for the NCRs based on the presence of all the six motifs (Table 4.1.3-1). Despite all the 185 NCR genes demonstrating the presence of one or the other six motifs, only 40 NCR genes out of all the 185 expressed genes were found to contain all the six motifs (Table 4.1.3-1). Motif 2 was present at its highest occurring four times in *Medtr3g069830*. One NCR gene *Medtr3g084910* was found to have high occurrence of motif 4, 1 and 5 with motif 4 and 5 occurring six times. Overall, motif 4, 1 and 5 as seen from alignment revealed stretch of longer element conservation. Except for one NCR gene *Medtr7g102170* that did not contain any of the six motifs, all the remaining NCRs (expressed and matched with the probe ID of our microarray data) were found to have at least one or the other combination of the six motifs.

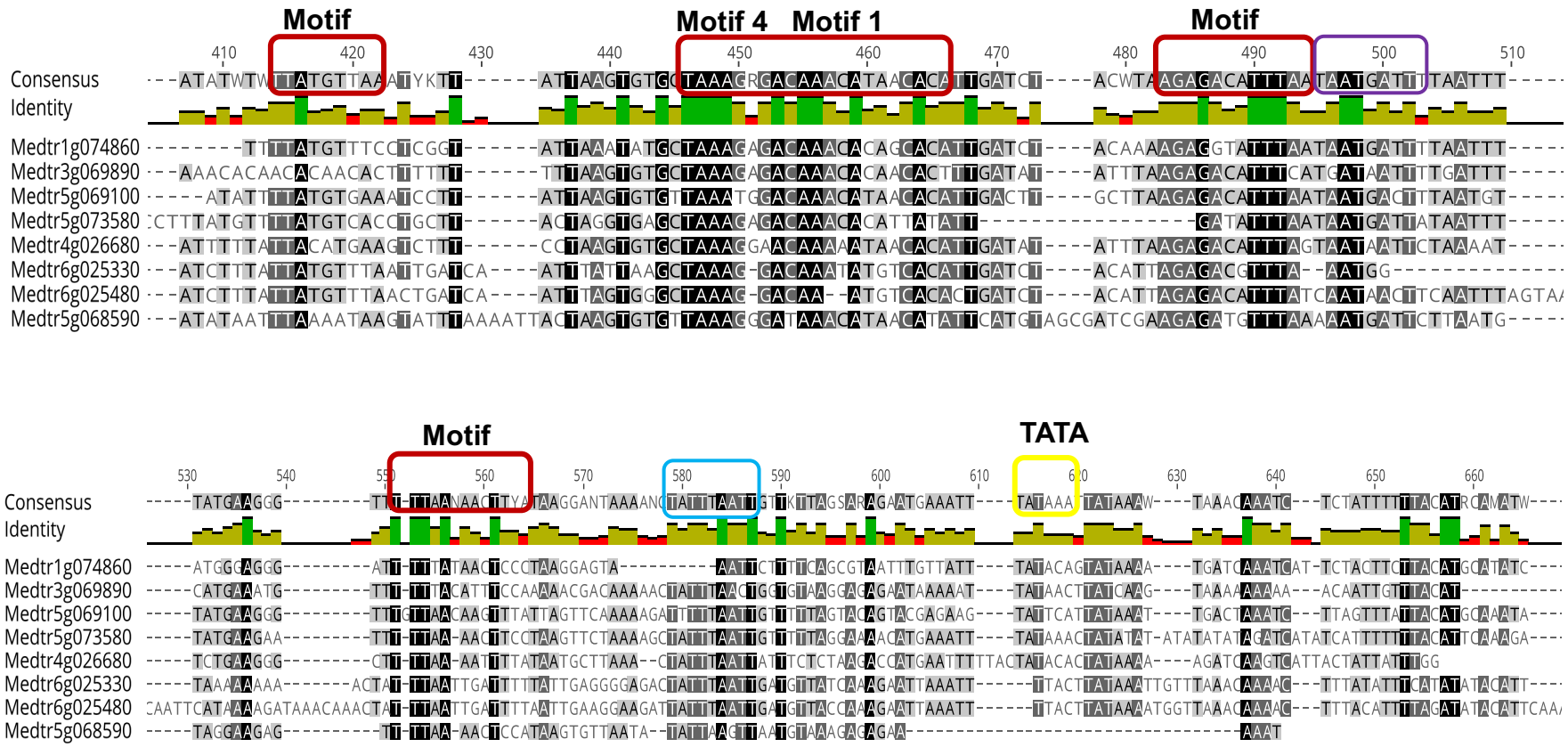


Figure 4.1.3: NCR promoter architecture analysis. Visualisation of Multiple Sequence alignment (MSA) using Genious in representative of eight NCR promoter regions [-1,-500] that showed conservation of the six motifs. The visualization was performed after aligning the sequence with EBI tools MAFFT (Multiple alignment using fast fourier transform). The sites of six motifs are shown in the red boxes, other conserved sites nearby the motif 2 region (shown in purple box) and motif 6 region (shown in purple boxes). Bars represent the identity of the conservation quality.

4.1.4 Potential TF(s) as regulators of NCR gene promoters

Transcription factors (TFs) play a major regulatory role in controlling gene expression by binding in a sequence specific manner (motifs) to the promoter region. We therefore searched for the TF(s) that could be regulators of the NCR genes by binding to the motifs in the NCR gene promoters. To investigate if the best conserved sites in our alignment and the MEME identified motif regions represent any previously known TF binding sites, the Tomtom tool (Gupta et al., 2007) in MEME-suite was used.

The scan with our PWM (described in our method section) served as our query and were tested against a number of databases: JASPAR plants 2016 (Mathelier et al., 2016), *Arabidopsis* PBM database (Franco-Zorrilla et al., 2014) and the CIS-BP database (Weirauch et al., 2014) for *Arabidopsis* (CISBP-At) and Medicago (CISBP-Mt) FZ (Figure 4.1.4). JASPAR plants database (Mathelier et al., 2016) is an open access database consisting of curated, non-redundant transcription factor (TF) binding profiles for multiple species in six taxonomic groups (vertebrates, nematodes, insects, fungi, urochordates and plants). The *Arabidopsis* PBM database (Franco-Zorrilla et al., 2014) utillises high-throughput protein binding microarrays (PBM) for the characterisation of the target sequence specificity of 63 plant specific TFs representing 25 families. The CIS-BP database (Weirauch et al., 2014) is catalogue of inferred sequence binding preferences developed by sampling DNA-binding domain (DBD) types from multiple eukaryotic clades. DNA binding sequence preferences for >1,000 TFs encompassing 54 different DBD classes from 131 diverse eukaryotes have been determined from known and inferred motifs sequence that coincide with ChIP-seq binding peak, are enriched in the promoter regions of diverse eukaryotes, and overlap eQTLs region in *Arabidopsis* (Weirauch et al., 2014).

Based on the sequence similarity, identity and E-values, eight TF targets were determined to be significant hits using an E-value <0.05. All the eight TFs correspond to transcription factor binding sites (TFBS) of *Arabidopsis* genes. Apart from the eight *Arabidopsis* TF(s), one likely binder protein called PEND (Plastid envelope DNA binding protein, figure 4.1.4 D) from *Pisum sativum* was found with hits outside the motif region at position 377-386 (JASPAR database). This PEND protein is thought to be involved in the binding of plastid nucleoids

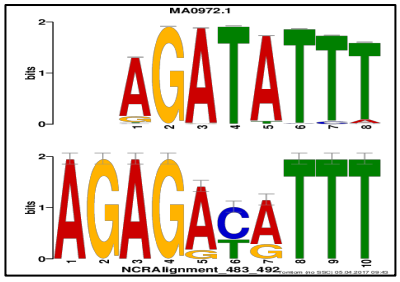
to the inner envelope membrane (Sato et al., 1998). We found eight putative TF(s) from the *Arabidopsis* database- CCA1, RVE1, ATHB15, ATHB16, AHL20, AHL25, WOX13 and AtGRP2B. Out of these eight potential TF(s), we have prioritized three TF(s) namely CCA1, RVE1 (Figure 4.1.4 A), ATHB15 and ATHB16 (Figure 4.1.4 B) because the target hits of alignment fell into our previously identified motifs 2 and 6. Other putative TF(s) of AHL family proteins (Figure 4.1.4 C) with potential binding sites to the conserved elements from alignment (Figure 4.1.3, highlighted with purple boxes) were also investigated. These AHL family proteins of TFs) were found to be strongly represented.

Amongst all the expressed 185 *NCR* genes, 134 *NCR* genes were found to have motif 2. Out of these 134 *NCR* genes with motif 2, there were 63 promoters in total that corresponded to our 12-mer motif 2 region with AGAGAC/TATTT. Altogether this corresponded to ~47% of the NCR genes containing an over-representation of the CCA1 binding site. Out of these NCR promoters, 15 NCR promoters contained the cis-regulatory element AGATATTT which was predicted to be bound by clock-related Myb TF proteins such as CCA1 (Manfield et al., 2007). Notably, we also found that the RVE1, a morning-phased TF that integrates clock and developmental pathways as a likely binder to this site represented (Figure 4.1.4 A). All the promoter of the *NCR* genes containing the exact *CAA1* binding sites of AGATATTT were differentially expressed with -1.5 < FC > 1.5 and p-value <0.05. On closer inspection of all the 185 expressed NCR genes, 27 promoter of the NCR genes were found to have the core element GATA (Table 4.1.4-1) that are also regulated by circadian rhythms (Reyes et al., 2004).

The other ATHBs transcription factor (ATHB15 and ATHB16) belonging to homeo-box protein family revealed a shared binding site with target consensus element of TAATWATT (Figure 4.1.3, in purple box, at alignment position of 495-502 region). This conserved element found from JASPAR plants 2016 database corresponded close to motif 2. Additionally, ATHB15 and ATHB16 were equally found to be a potential binder in the region overlapping our motif 6 (Figure 4.1.3) from CISBP-At database. ATHB16 function as a growth regulator in response to photoperiodism for development pathway while ATHB15 is a transcription factor regulating meristem and lateral organ

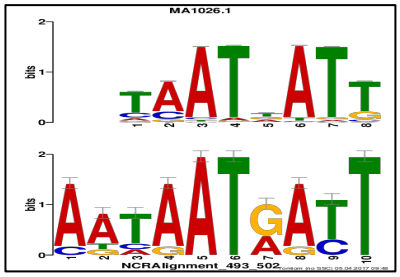
development (Ochando et al., 2006; Wang et al., 2003b). Other putative binders include AT-hook Motif Nuclear Localised, AHL protein family (AHL20 and AHL25) targeting the well conserved site AWTTAATT closely following our motif 6 region (Figure 4.1.3 in blue box and Figure 4.1.4c). AHL20 is involved in the negative regulation of plant defenses while AHL25 is required for negative regulation of GA (Lu et al., 2010; Matsushita et al., 2007). Considering the conserved sites for TF search from the sequence alignment, in general ATHB and AHL protein family were highly represented as significant hits (data not shown).

(A)



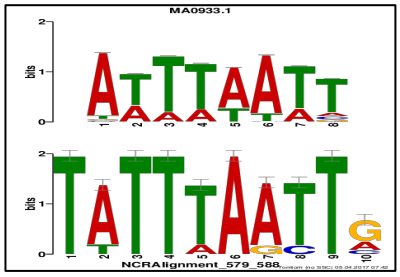
CCA1, RVE1 → motif2 (Circadian Clock Associated)
E-value: 0.0152058
CCA1 ortholog annotated as Late elongated hypocotyl-like protein
RVE1 ortholog annotated as Myb transcription factor

(B)



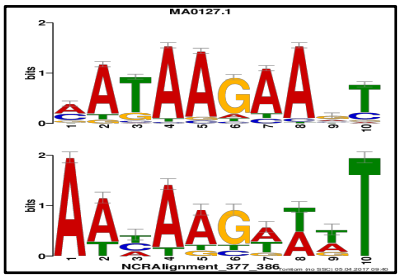
ATHB-15, ATHB-16 → motif6 (Plant development)
E-value: 0.0152808
Medicago ortholog ATHB15 annotated as class III homeodomain leucine zipper protein
Medicago ortholog ATHB16 homeobox associated leucine zipper protein

(C)



AHL20, AHL25 → Outside the motif region in promoter
E-value: 0.00229149

(D)



PEND → Outside the motif region in

Figure 4.1.4: Putative transcription factor hits from *A.thaliana* database identified. Known TF from TF-target databases aligned with the motif region and the conserved sites from multiple sequence alignment using Tomtom scan in MEME suite. The target region and the query are derived from database JASPAR Plants 2016 and Franco-Zorrila Arabidopsis PBM. (A) CCA1 and RVE1 binding site (AGATATTT) alignment with query motif 2 (AGAGACATTT). (B) ATHB15 (TAATTATT) and ATHB16 (TAATAATT) binding site target alignment (C) AHL20/AHL25 binding target site (TATTTAATTG) alignment in 579-588 conserved region from MSA. (D) PEND protein binding site at the alignment region of 377-386 from MSA.

Table 4.1.4-1: Promoters of the 27 *NCR* genes with putative CCA1 binding sites that contained the core GATA element. Occurrence site, DNA strand and p-values as obtained from FIMO scan in MEME suite.

Gene ID	Start	Stop	Strand	p-value	Element
Medtr5g070410	346	357	+	2.25E-07	AGA**GATA**TTTAA
Medtr5g061400	349	360	+	2.25E-07	AGA**GATA**TTTAA
Medtr3g069830	379	390	+	2.25E-07	AGA**GATA**TTTAA
Medtr6g006350	409	420	+	2.25E-07	AGA**GATA**TTTAA
Medtr6g006240	422	433	+	2.25E-07	AGA**GATA**TTTAA
Medtr5g072420	457	468	+	1.54E-06	AGA**GATA**TTTAT
Medtr5g072310	354	365	+	1.91E-06	AGA**GATA**ATTAA
Medtr5g037780	353	364	+	3.72E-06	AAA**GATA**TTTAA
Medtr3g031340	368	379	+	3.72E-06	AAA**GATA**TTTAA
Medtr3g033100	390	401	+	3.72E-06	AAA**GATA**TTTAA
Medtr3g033260	401	412	+	3.72E-06	AAA**GATA**TTTAA
Medtr1g046020	472	483	+	6.69E-06	AGA**GATA**AATAA
Medtr1g058880	278	289	+	2.63E-05	AGA**GATA**TTTAC
Medtr3g069870	396	407	+	2.63E-05	AGA**GATA**TTTAG
Medtr5g071880	348	359	+	2.95E-05	AGA**GATA**TTTTA
Medtr5g057460	433	444	+	2.95E-05	AGA**GATA**TTTTA
Medtr4g033830	62	73	+	4.78E-05	AAA**GATA**AATAA
Medtr4g031430	197	208	+	4.78E-05	AAA**GATA**AATAA
Medtr5g063600	215	226	+	4.78E-05	AAA**GATA**AATAA
Medtr2g045290	233	244	+	4.78E-05	AAA**GATA**AATAA
Medtr4g031520	240	251	+	4.78E-05	AAA**GATA**AATAA
Medtr1g044500	142	153	-	4.78E-05	AAA**GATA**AATAA
Medtr3g020930	405	416	+	7.26E-05	AAA**GATA**TACAA
Medtr5g059740	489	500	+	7.83E-05	AGA**GATA**CATAA
Medtr5g066750	256	267	-	8.46E-05	AGA**GATA**TTTTT
Medtr4g031380	325	336	+	9.46E-05	AAAGA**GATA**TAT
Medtr7g010200	330	341	+	9.46E-05	AAAGA**GATA**TAT

Table 4.1.4-2: Promoters of the 36 *NCR* genes with putative CCA1 binding sites that contained the AGACATTT element with their occurrence site, DNA strand and p-values as obtained from FIMO scan in MEME suite.

Gene ID	Start	Stop	Strand	p-value	Element
Medtr4g060610	331	342	+	0.000000102	AGAGACATTTAA
Medtr7g028550	340	351	+	0.000000102	AGAGACATTTAA
Medtr5g026070	343	354	+	0.000000102	AGAGACATTTAA
Medtr3g062830	349	360	+	0.000000102	AGAGACATTTAA
Medtr6g060320	351	362	+	0.000000102	AGAGACATTTAA
Medtr7g045410	354	365	+	0.000000102	AGAGACATTTAA
Medtr2g063470	360	371	+	0.000000102	AGAGACATTTAA
Medtr7g045910	363	374	+	0.000000102	AGAGACATTTAA
Medtr5g061120	364	375	+	0.000000102	AGAGACATTTAA
Medtr7g029760	368	379	+	0.000000102	AGAGACATTTAA
Medtr1g042910	373	384	+	0.000000102	AGAGACATTTAA
Medtr1g042200	374	385	+	0.000000102	AGAGACATTTAA
Medtr1g042940	377	388	+	0.000000102	AGAGACATTTAA
Medtr5g069100	380	391	+	0.000000102	AGAGACATTTAA
Medtr4g057120	388	399	+	0.000000102	AGAGACATTTAA
Medtr1g042850	394	405	+	0.000000102	AGAGACATTTAA
Medtr5g063490	432	443	+	0.000000102	AGAGACATTTAA
Medtr5g063890	433	444	+	0.000000102	AGAGACATTTAA
Medtr5g063780	438	449	+	0.000000102	AGAGACATTTAA
Medtr5g063580	477	488	+	0.000000102	AGAGACATTTAA
Medtr7g016440	339	350	+	0.000000428	AGAGACATTTAT
Medtr5g059670	351	362	+	0.000000428	AGAGACATTTAT
Medtr7g008940	355	366	+	0.000000428	AGAGACATTTAT
Medtr5g056360	376	387	+	0.000000428	AGAGACATTTAT
Medtr3g015940	204	215	-	0.00000179	AAAGACATTTAA
Medtr3g033930	248	259	+	0.00000179	AAAGACATTTAA
Medtr2g060960	316	327	+	0.00000179	AAAGACATTTAA

Table 4.1.4-2 (continued):

Gene ID	Start	Stop	Strand	p-value	Element
Medtr5g061160	406	417	+	0.00000179	AAAGACATTTAA
Medtr4g026750	344	355	+	0.0000125	AGAGACATTTAG
Medtr7g008970	347	358	+	0.0000125	AGAGACATTTAC
Medtr7g071310	356	367	+	0.0000125	AGAGACATTTAG
Medtr4g026680	365	376	+	0.0000125	AGAGACATTTAG
Medtr2g104570	417	428	+	0.0000154	AGAGACATTTGA
Medtr6g044730	30	41	-	0.0000411	AGAGACATTTGT
Medtr4g060720	346	357	+	0.0000919	TAAGACATTTAA
Medtr7g071220	383	394	+	0.0000919	AAAGACATTTTA

4.2 Discussion

Differential gene expression analysis from our transcriptomic data and investigation of DE gene promoters led to the identification of putative motif sites in the promoters of genes responsive to N and rhizobia. Consistent with the report of NCR gene expression only in symbiotic nodules, our microarray data found this and also found co-regulation of NCR genes with nodulation pathway genes, we sought to investigate the promoter region for presence of regulatory binding sites. Work in this chapter showed the presence of conserved motifs from our *de novo* motif discovery in co-expressed promoter regions. As well as the six motifs (TCATGAAAGGTT, TATAAAGTGATCA, CAACACATTGAT, AGAGACATTTAA, TTTTACAACTCC) found from MEME analysis, conserved stretches of these elements were characterised (Figure 4.1.3). As seen in the table 4.1.2-1 the majority of promoter with the conserved motifs were found to consist of nodulins and NCR encoding genes.

Promoters of NCRs with significant motifs as PSSMs could serve as putative regulatory binding sites. This led to the hypothesis that if these motif sites were conserved, perhaps it is a feature of NCR upstream regions where transcription factors are recruited for their regulation. The fact that all of the NCR genes were significantly upregulated (FC>1.5, p-value <0.05) in the

presence of rhizobium along with genes involved in nodulation pathway also suggested the existence of a co-regulatory mechanism with nodulins.

Promoter sequence diversity is evident from the wide variation in motif occurrence as shown in Table 4.1.3-1. There seems to be no particular pattern in motif occurrence across promoters, except for motif 4,1 and 5 seemingly indicating a stretch of a longer motif. We are not the first to report the presence of motifs in the NCR upstream region, but our data is in line with previous work. The 12 bp of the six motifs we found overlapped the longer 41-50bp motifs in the NCR promoter as previously reported by (Nallu et al., 2013). Nallu et al (2013) found that 566 NCRs were expressed over the stages of nodule formation, Nallu et al (2013) used a custom Affymetrix microarray with probes for 684 Medicago *DEFLs* (defensin like family genes) method enabled them to analyse more NCRs than on our NimbleGen microarray. While our results are consistent with NCR expression being dependent on the number and volume of the rhizobia in nodules, we did not find any of the other motifs that Nallu et al (2013) found ID1 binding site, Auxin Response Factor (ARF) binding site, Dof protein binding site, and MADS box gene binding sites) over represented in our differentially expressed *NCR* genes.

Understanding gene regulation and expression often poses challenges due to the degeneracy of transcription factor binding sequences. The over-representation of clock related binding sites (AGAC/TATTT) in the upstream region of NCR genes suggests a novel regulation of NCR expression under circadian control, although the exact contribution remains unclear. Previous *in vitro* assay characterisation identified direct binding of CCA1 and LHY to the evening element motif AAATATCT that was over-represented in the promoters of circadian regulated gene that show a peak expression in the evening (Edwards et al., 2005; Harmer et al., 2000; Michael et al., 2008).

Circadian oscillation of gene regulation plays a vital role as an internal time-keeper to align optimum growth, development and assimilation. Independent analysis of the root and shoot circadian rhythm in *Arabidopsis* has shown that the endogenous root clock exhibits an extreme version of organ specificity (James et al., 2008). The study reported only a subset of 3.2% of root genes that display rhythmicity, including the *MYB* gene *RVE1,* have their

peak expression delayed in roots relative to shoot (James et al., 2008). The organ-specific circadian clock was further defined in another study that evaluated the differences of light input effect in roots and shoots (Bordage et al., 2016). This indicates that the circadian clock depends on the organ type and environment. This is in line with our finding of RVE1 binding sites in the NCR upstream region as NCRs are known to have extreme root nodule-specific expression.

It has been shown that all plant hormone signalling pathways that influence growth and development have an enrichment of circadian-responsive genes (Covington et al., 2008). One of the pieces of evidence implicating the circadian clock in root development is the finding that it is re-phased during LR development in co-ordination with auxin signalling (Voss et al., 2015). The study demonstrated that the circadian clock is required for LR emergence and development to gate auxin responses in root primordial cells. Similarly the clock and hormones regulate nodulation in legume roots. The involvement of the phytohormone cytokinin on regulation of nodule organogenesis has been elaborated with the identification of two TFs *NSP2* and *bHLH476* as primary targets (Ariel et al., 2012). Moreover, altered cytokinin response affecting the diurnal rhythm of circadian oscillations in *lhy* and *cca1* mutants of *Arabidopsis* have been studied (Zheng et al., 2006). These findings highlight the influence of phytohormones in concert with circadian regulation of nodule organogenesis.

Another piece of evidence that the circadian system integrates with phytohormonal pathway comes from the study that *RVE1* affects hypocotyl elongation and increases auxin levels in Arabidopsis (Rawat et al., 2009). *RVE1* is a myb-like TF homologous to core circadian genes *CCA1*, *LHY*. These circadian genes *CCA1* and *LHY* (represented by single amino acid sequence in *M. truncatula* and *G. max*; (Hecht et al., 2005)) with myb-like domain contain a distinct motif SHAQKYF that binds to the EE motif (Schaffer et al., 1998; Wang and Tobin, 1998). Therefore, the over-representation of CCA1/RVE1 binding sites on *NCR* genes which expresses only in nodules, the accumulating evidence of circadian oscillations to be organ-specific, along with phythormonal integration, strongly suggest the possible existence of a nodule circadian clock.

Additional clock-controlled regulation that is important for understanding rhizobial colonisation in Medicago roots is the expression of defense response genes against pathogen invasion. In a study of plant defence, the Arabidopsis *Pseudomonas syringae* and insect immune response-regulating R-mediated genes were found to be under circadian control (Bhardwaj et al., 2011; Goodspeed et al., 2012; Shin et al., 2012). Significant enrichment of CCA1 binding sites were also reported in the promoters of genes that are induced upon invasion of the biotrophic pathogen *Hyaloperonospora arabidopsidis* (Goodspeed et al., 2012; Zhang et al., 2013). With the postulated role of NCR expression as a defence response, induced by rhizobium, the enrichment of CCA1-binding sites in NCR promoters could enable reception of a signal in IRLC legumes such as *M. truncatula* to prepare for 'symbiont infection' and nodule development.

Following the hypothesis that gene duplication can enable the gain of new function in the course of evolution (Flagel and Wendel, 2009), with evidence that the CCA1 target site underwent a massive genomic expansion (Nagel et al., 2015) and with myriad of biological functions under the circadian rhythm, we propose that NCRs have a role beyond just a defense mechanism.

In this study, we proposed that members of the large NCR gene family appear to have additional roles under the regulation of our identified TFs. We speculate that the presence of likely binders of the ATHB and AHL protein family, Myb TFs like CCA1⁄LHY and RVE1 is indicative of indel (insertions and deletions) events leading to the recruitment of transcription factors contributing to the regulation of endosymbiotic nitrogen fixation. The putative transcription factors belonging to superfamily of ATHB and AHL protein family indicate the possibility of NCRs function in plant development beyond defense. It is possible that the *NCR* genes have co-evolved with another ancient gene family for gain of new symbiotic function through gene duplications that may have occurred in Arabidopsis and IRLC legumes. This supports a likely existence of likely multiple functions of NCRs as DEFLs have been reported to play dual roles in defense and developmental plant signalling during plant-microbial interaction. As such, functional characterization of NCRs and how they are regulated will help in elucidating any additional roles of NCRs.

Chapter 5: Identification of *Medicago truncatula* transcription factors as putative NCR regulators

5.1 Introduction

As described in Chapter 4, interaction of cis-acting TF with DNA is highly sequence specific. To identify promoter binding sites we used a TOMTOM scan from the TF database of the commonly studied non-legume model plant *Arabidopsis thaliana*. This was also used to compare the *de novo* motifs identified in NCR promoters to TF(s) that were biochemically characterized or previously identified. Due to the degree of evolutionary conservation between plants, the functional features of experimentally characterized TF genes in *Arabidopsis* are likely to be comparable to *Medicago* (Van Bel et al., 2018). We followed the hypothesis that functional features for TF gene regulation (the cis regulatory elements) are shared across species and that TF regulators would also be homologous, as orthologous genes tend to biologically conserve their function in different organisms during evolutionary diversification.

We used comparative genomic methods to identify gene-homology relationships between *Arabidopsis* TF(s) and Medicago and ask if putative TF ortholog genes binding the motifs that we identified exist in the Medicago genome. We employed a reciprocal blast method to query the amino acid sequence of *Arabidopsis* proteins predicted to bind the motifs against Medicago genome and thus identify orthologs.

To identify Medicago mutants defective in putative TF orthologs the Medicago database of mutants developed by Tnt1 retrotransposon insertion and Fast Neutron Bombardment (FNB) was used (Sun et al., 2018; Tadege et al., 2008). A genotyping screen to identify mutants that were homozygous was carried out, then the level of mRNA expression determined to ask if the mutants were defective in expression of the genes carrying a mutation.

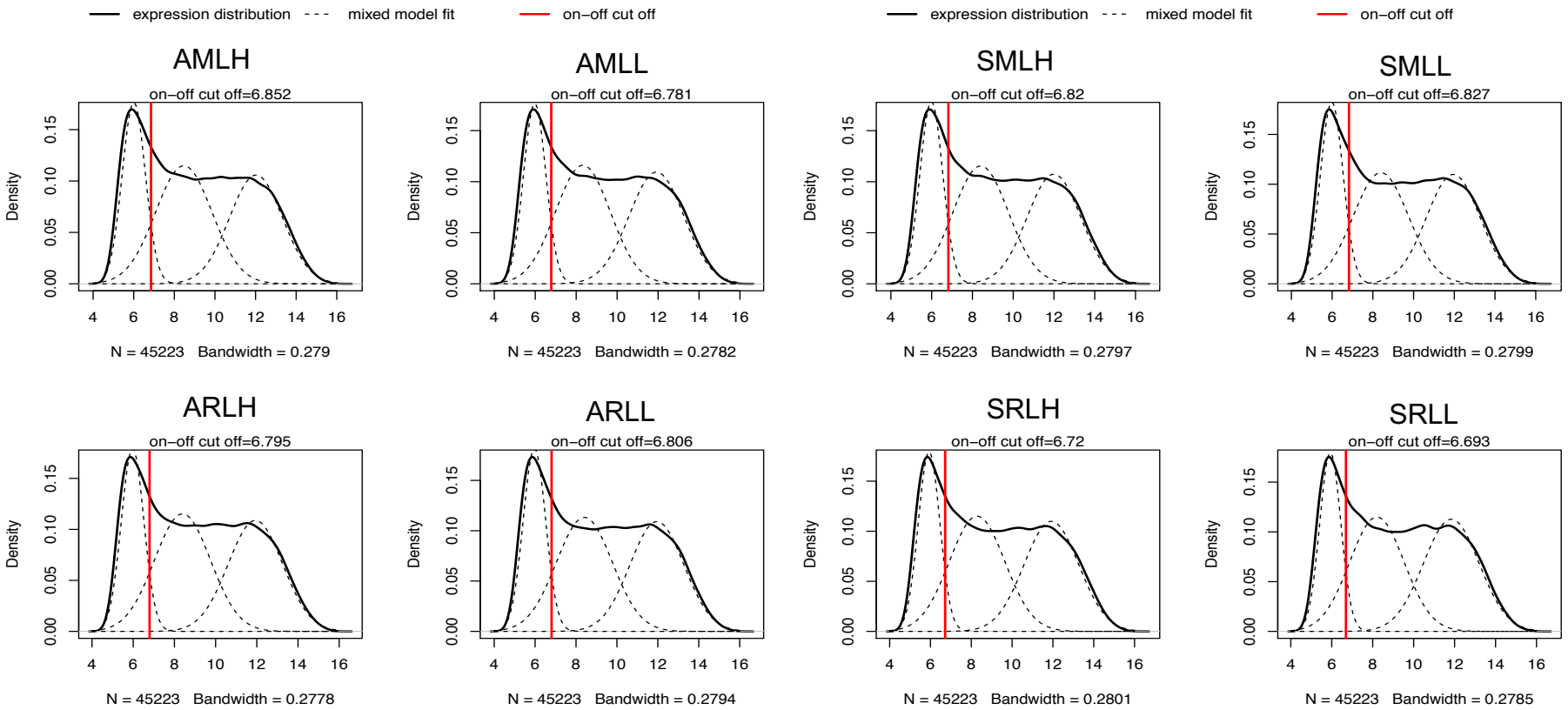


Figure 5.1.1-1: Normal distribution fit to the Log2 gene expression density profiles of transcriptomic data for classification of distinct expression mode using MClust R package. X-axis represent the kernel density and Y-axis represents the log2 expression signal of microarray hybridization. The threshold values for the treatment sample described are given in Table 3.1.3-1. The threshold expression boundary was determined using a mixture model. The black solid smooth curve represents the gene expression distribution pattern analysed from the array probe set. The red line denotes threshold level for the genes to be considered “on-off”.

5.1.1 Medicago orthologs from putative TF(s) and their expression

Using a reciprocal protein BLAST with *Arabidopsis* TF, eight TF(s) (CCA1, RVE1, ATHB15, ATHB16, AHL20, AHL25, WOX13, AtGRP2B) were queried against the Medicago genome 4.0 version. A total number of 31 unique Medicago orthologous genes were identified from the search; the *Arabidopsis* gene loci for the TF(s) and number of Medicago orthologs expressed are given in Table 5.1.1-1.

Based on the E-value score, identity score and alignment score, 21 *Medicago truncatula* orthologs were selected for study (Table 5.1.1-1). This included seven *Arabidopsis* gene loci with multiple *Medicago* orthologs and one *Arabidopsis* gene (*AtGRP2B*) with one ortholog. For each *Medicago* ortholog we asked if it was expressed within root sample mRNA. To understand the global spread of A17 and *sunn-1* gene expression distributions kernel density estimate (Figure 5.1.1-1) was used to determine that 17 Medicago orthologs were expressed above threshold level in both the genotypes. The threshold level for A17 data was in the range of 6-7-6.8 (log2 gene expression) while for *sunn-1* the threshold level was in the range of 6.6-6.8 (log2 gene expression) (Figure 5.1.1-1). The threshold level of expression from the kernel density estimate determined which genes were considered to be expressed.

Amongst the 17 *M. truncatula* orthologs was *Medtr7g118330* (*late elongated hypocotyl-like protein, LHY*), a putative *CCA1* ortholog, and *Medtr3g064500* and *Medtr5g076960* (MYB transcription factor proteins) as putative *RVE1* orthologs (Table 5.1.1-1). Further evaluation of our putative *CCA1* and *RVE1* orthologs for their transcript abundance within our expression dataset show expression of the one *CCA1* ortholog and the two *RVE1* ortholog genes (Figure 5.1.1-2) in all the treatment conditions for both the wild type and mutant background (A17 and *sunn-1*), except for the *RVE1* ortholog *Medtr3g064500*. The transcript profile of CCA1 and RVE1 in the wild-type A17, irrespective of rhizobium absence or presence, follows a similar trend, with lower expression for plants grown in N deplete conditions than in N-replete conditions (Figure 5.1.1-2). In the *sunn-1* mutant we observed an increased

expression of CCA1 and RVE1 Medicago ortholog transcripts in rhizobium-inoculated seedlings under deplete N conditions, compared to the same conditions in A17 wild-type (Figure 5.1.1-2).

Table 5.1.1-1 Medicago orthologs that are expressed in transcriptomic data from root samples. There are 17 non-redundant unique Medicago orthologs expressed on our data.

Gene name	Query (*Arabidopsis* gene locus)	*A. thaliana* gene annotation	*M. truncatula* ortholog(s) gene locus	*M. truncatula* gene annotation	Identity (%)	E-value
CCA1	AT2G46830.1	Circadian clock associated protein 1	Medtr7g118330.1	late elongated hypocotyl-like protein	51.293	1.51E-49
RVE1	AT5G17300.1	REVEILLE 1, Myb-like transcription factor	Medtr5g076960.1	myb transcription factor	46.847	2.67E-58
ATHB15	AT1G52150.2	Class III homeodomain leucine zipper protein	Medtr2g030130.1 Medtr2g094520.1 Medtr2g101190.1 Medtr3g109800.1 Medtr8g013980.1	class III homeodomain leucine zipper protein	66.348 61.437 84.982 78.791 83.552	0 0 0 0 0
*ATHB*16	AT4G40060.1	homeodomain leucine zipper class I (HD-Zip I) protein	Medtr3g086790.1	homeobox associated leucine zipper protein	44.615	3.18E-66
AHL20	AT4G14465.1	AT hook motif nuclear localized protein	Medtr5g080580.1	DUF296 domain protein	67.832	6.01E-53

Table 5.1.1-1 (continued) Medicago orthologs that are expressed in transcriptomic data from root samples. Redundant orthologous genes in AHL20 and AHL25 family have been underlined.

Gene name	Query (*Arabidopsis* gene locus)	*A. thaliana* gene annotation	*M. truncatula* ortholog(s) gene locus	*M. truncatula* gene annotation	Identity (%)	E-value
AHL20	AT4G14465.1	AT hook motif nuclear localized protein	Medtr1g079810.1 Medtr3g100470.1 Medtr4g098450.1 Medtr5g011520.1 Medtr5g091630.1 Medtr7g080980.1	AT hook motif DNA-binding family protein	51.582 49.495 62.147 65.409 52.83 51.22	1.13E-79
AHL25	AT4G35390.1	AT hook protein of GA feedback 1	Medtr1g079810.1 Medtr4g098450.1 Medtr5g011520.1 Medtr5g091630.1 Medtr7g080980.1	AT hook motif DNA-binding family protein	64.557 53.509 70.44 63.03 53.211	1.65E-59 4.15E-61 1.55E-59 1.21E-63 3.00E-58
AtGRP2B	AT4G13850.1	Glycine rich RNA binding protein	Medtr3g084040.1	RNA-binding (RRM/RBD/RNP motif) family protein	65.812	8.24E-49

Table 5.1.1-2 Medicago orthologs that are not expressed in transcriptomic data from root samples. Redundant orthologous genes in *AHL20* and *AHL25* family have been underlined.

Gene name	Query (*Arabidopsis* gene locus)	*A. thaliana* gene annotation	*M. truncatula* ortholog(s) gene locus	*M. truncatula* gene annotation	Identity (%)	E-value
AHL20	AT4G14465.1	AT hook motif nuclear localized protein	Medtr1g073860.1 Medtr1g044155.1 Medtr3g068035.1 Medtr8g036060.1	AT hook motif DNA-binding family protein	59.933 58.654 60.829 69.481	9.9E-92 1.77E-92 2.12E-68 3.4E-59
AHL25	AT4G35390.1	AT hook protein of GA feedback 1	Medtr1g073860.1 Medtr1g044155.1 Medtr3g068035.1 Medtr8g036060.1	AT hook motif DNA-binding family protein	68.421 71.053 60.829 59.307	1.35E-59 5.22E-61 2.12E-68 4.5E-64
ATHB15	AT1G52150.2	class III homeodomain leucine zipper protein	Medtr3g109800.2 Medtr4g058970.1	class III homeodomain leucine zipper protein	77.778 66.865	0 0
ATHB16	AT4G40060.1	homeodomain leucine zipper class I (HD-Zip I)	Medtr8g089895.1	homeobox associated leucine zipper protein	44.277	1.04E-69
WOX13	AT4G35550.1	WUSCHEL-related homeobox gene family	Medtr1g115315.1 Medtr1g115315.2 Medtr3g115620.1	wuschel-related homeobox protein	50.671 53.759 52.991	50.671 53.759 52.991
RVE1	AT5G17300.1	REVEILLE 1, Myb-like transcription factor	Medtr3g064500.1	myb transcription factor	47.945	3.11E-64

Of the six Medicago orthologs identified for putative developmental TF proteins, ATHB15 (*Medtr3g086790*) and ATHB16 (five orthologs, *Medtr2g030130*, *Medtr2g101190*, *Medtr3g109800*, *Medtr8g013980*) were found to have similar expression pattern compared to one ortholog (*Medtr2g094520*) with lower expression (Figure 5.1.1-2). AHL20 and AHL25 with redundant ortholog genes (*Medtr1g079810*, *Medtr4g098450*, *Medtr5g011520*, *Medtr5g091630*, *Medtr7g080980*) represented the TF family of AHLs (Table 5.1.1-2). The *Medicago* AHL orthologs were found to share similar expression profile in A17 and *sunn-1* except for one orthologous gene *Medtr5g091630* (Figure 5.1.1-2). In general, the ATHB and AHL sequences were found to have a large number of putative hits in the Medicago genome and these were found to be redundant at the sequence level. They were also found to be largely redundant at the expression level (Table 5.1.1-2) (although the location of expression was not analysed

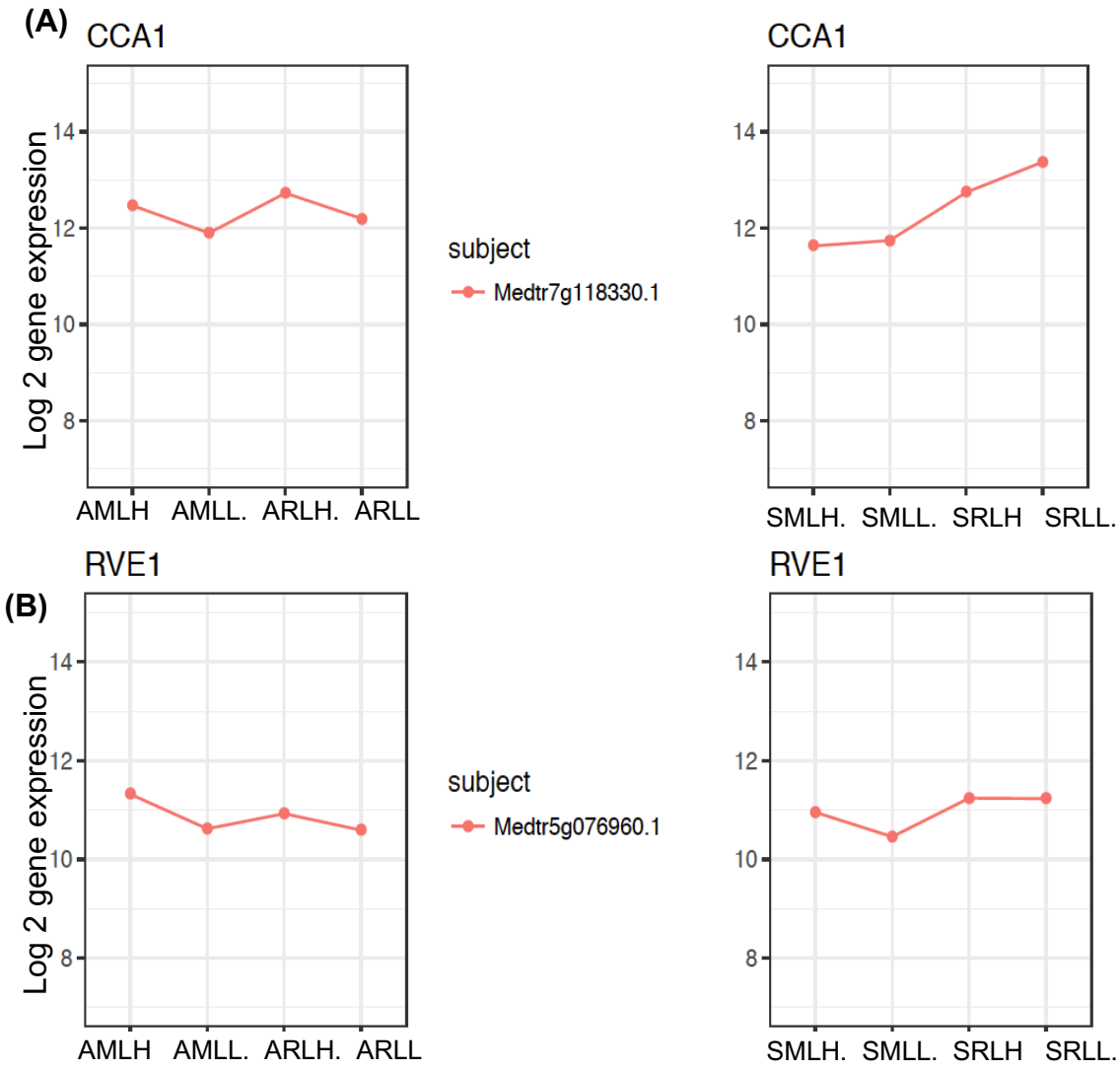


Figure 5.1.1-2: Transcript profile of *MYB* gene family orthologs of CCA1 and RVE1 in deplete or replete N condition (+rhizobia/-rhizobia) conditions in A17 and *sunn-1*. (A) Top left shows log2 expression of CCA1 ortholog *Medtr7g118330* in A17 in mock-replete N (AMLH), mock-deplete N (AMLL), rhizobia-replete N (ARLH) and rhizobia-deplete N (ARLL) conditions; top right shows expression in the *sunn-1* mutant in mock-replete N (SMLH), mock-deplete N (SMLL), rhizobia-replete N (SRLH) and rhizobia-deplete N (SRLL) conditions. (B) Bottom left shows log2 expression of RVE1 ortholog *Medtr5g76960* in A17 in AMLH, AMLL, ARLH and ARLL conditions while bottom right shows expression in *sunn-1* in SMLH, SMLL, SRLH and SRLL conditions.

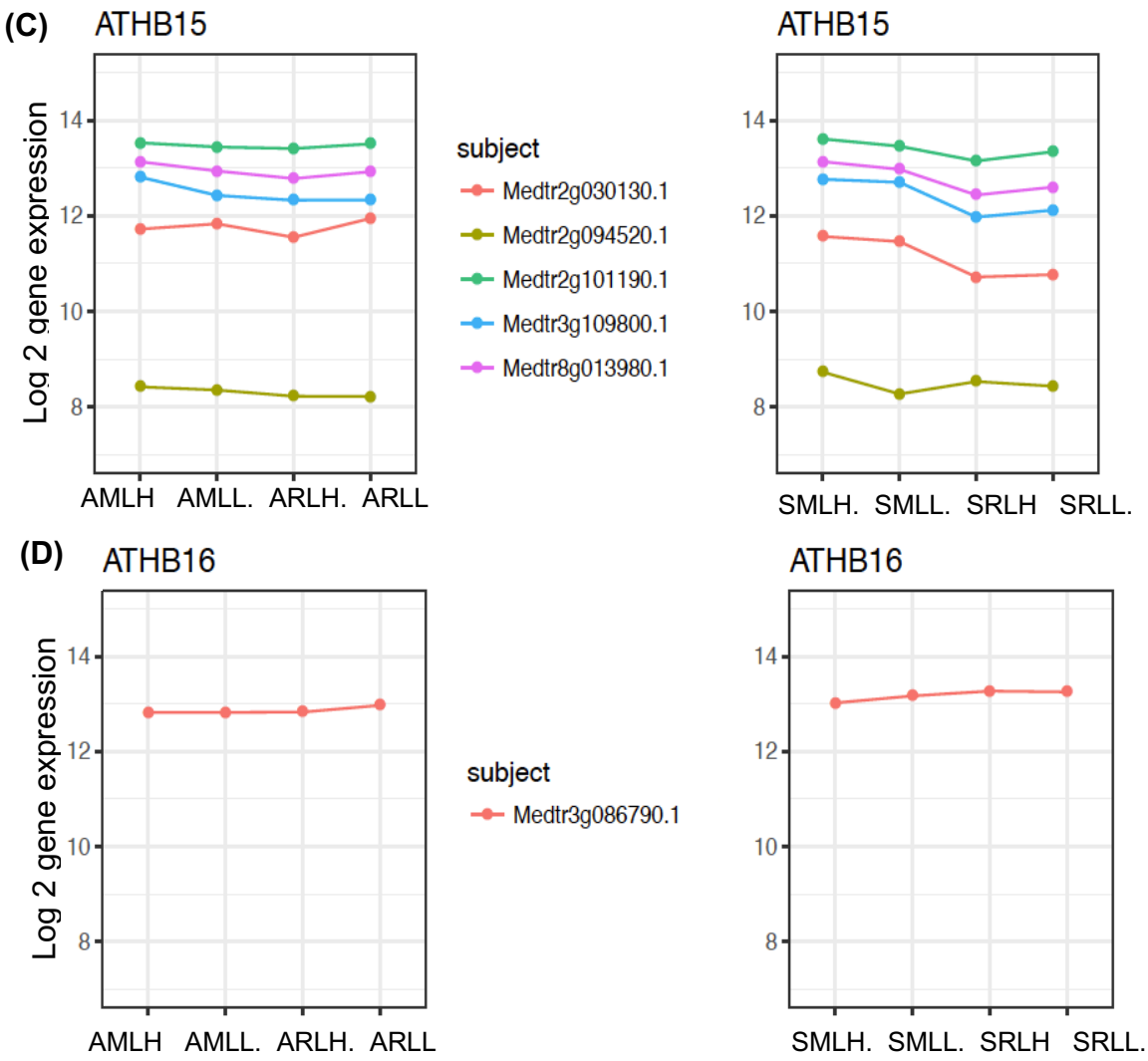


Figure 5.1.1-2 (continued) Transcript profile of homeobox gene family orthologs (C) *ATHB15* and (D) *ATHB16* in deplete or replete N condition (+rhizobia/-rhizobia) conditions in A17 and *sunn-1*.

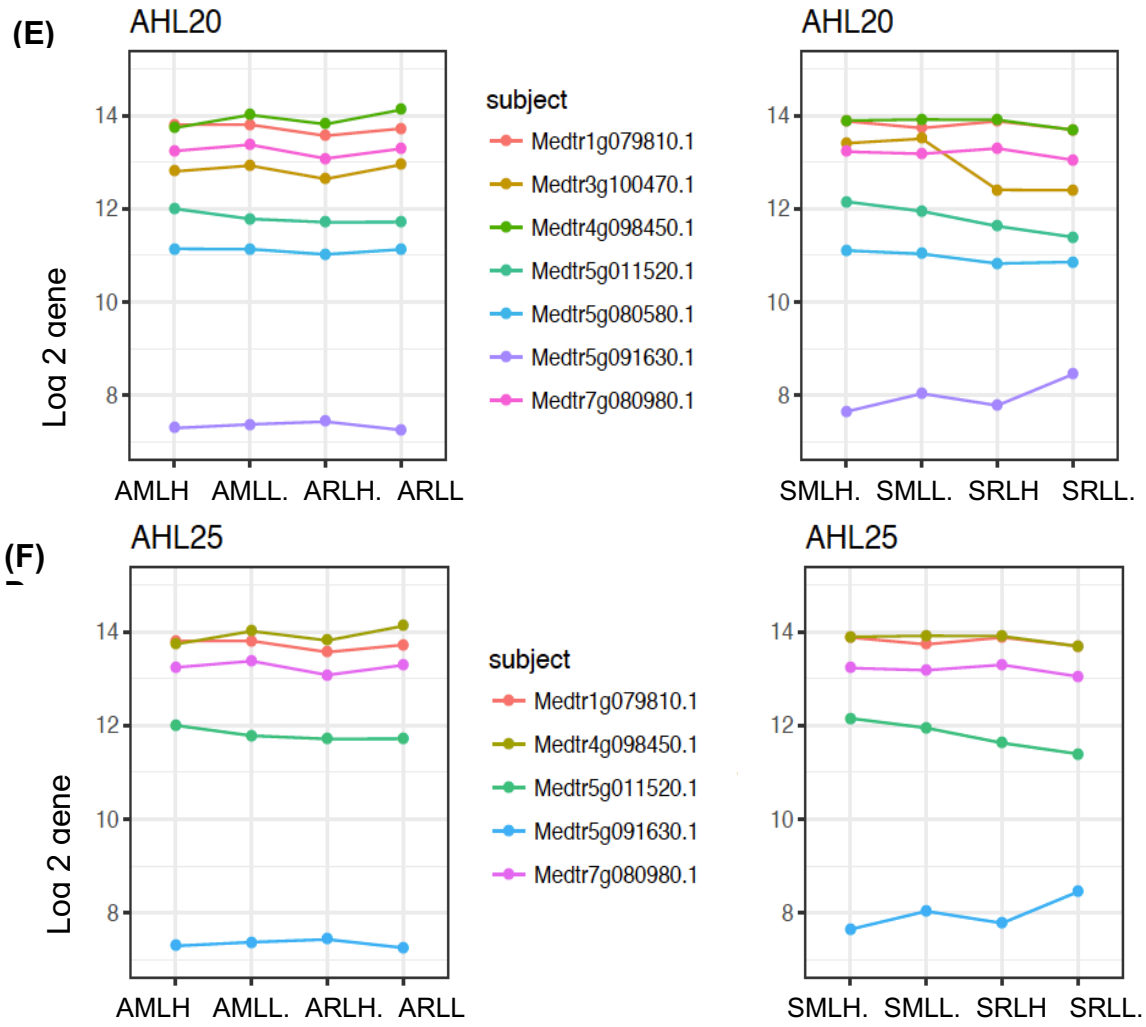


Figure 5.1.1-2 (continued) Transcript profile of *AHL20* (E) and *AHL25* (F) deplete or replete N condition (+rhizobia/-rhizobia) conditions in A17 and *sunn-1*.

5.1.2 Identification of *Tnt1* Medicago mutants

To elucidate the function of the putative NCR regulators we sought to search for mutants that were disrupted in the putative orthologous TF genes. The *M. truncatula* mutant population generated by tobacco retrotransposon (*Tnt1*) insertion developed at the Samuel Roberts Noble Foundation (Sun et al., 2018; Tadege et al., 2008) offers the most comprehensive catalog of putative mutants and thus was the source for this work.

Tnt1-tagged Medicago mutants have already been utilised for forward genetics screening in a number of laboratories e.g (Ameline-Torregrosa et al., 2008; Benlloch et al., 2006; Tadege et al., 2008). In order to identify the Medicago mutant lines, the nucleotide sequence of ortholog genes was used as the query sequence for BLAST search. A *M. truncatula* mutant database (Sun et al., 2018; Tadege et al., 2008), at the Noble Research Institute holds a collection of Medicago mutant lines developed in the R108 background by Tnt1 retrotransposon insertion. The *Tnt1* Flanking Sequence Tag (FST) flanks the insertion sites.

Since the Medicago mutant lines are in the R108 background, so first all ortholog gene sequences were identified in the R108 genome (see Methods). A blastn search with the indexed gene sequence of the 7 TF proteins (clock related - CCA1, RVE1; development related - ATHB15, ATHB16, AHL20, AHL25 and WOX13) from R108 background resulted in identification of 37 high-confidence *Tnt1*-tagged mutant lines with FST insertions (Table 5.1.2-1). These mutant lines were then prioritised based on the most significant E-values, their alignment identity >95% and known FST sites for the insertions.

For the CCA1 ortholog gene *Medtr7g118330* four mutant lines were identified with FST insertions in 6th exon (NF2784, NF17115), upstream region (NF20691) and 2nd intron (NF16461). For the RVE1 ortholog gene *Medtr5g076960* only one mutant NF13921, with a FST insert on the 2nd exon, was found. For the orthologous genes of ATHBs (*Medtr2g094520*- NF1297; *Medtr3g109800*- NF5993, NF17550; *Medtr2g030130*- NF0908, NF1084, NF1729; *Medtr4g058970*- NF5999; *Medtr8g013980*- NF11854) there were nine mutant lines (Table 5.1.2-1). For the *AHL* gene family (with redundant

orthologous genes) there were 20 mutant lines and the WOX13 family of TF there were two mutant lines (Table 5.1.2-1). A search for mutants was also carried out in the Medicago population generated by FNB mutagenesis. However, it resulted in poor alignments and the E-values were not significant as compared to *Tnt1* BLAST results (data not shown).

Table 5.1.2-1: Tnt1 mutant lines identified from the blastn search from the Medicago mutant database (Noble Research Foundation). Medicago genes corresponding to orthologous *Arabidopsis* TF(s), FST start and end site co-ordinates with the respective mutant lines found for each gene are given below. Genes in this table were expressed in our microarray data. Mutant line NF5993 had two hits for the two transcript variants Medtr3g109800.1 (expressed) and Medtr3g109800.2 (not expressed, see the next Table 5.1.2-2)

Medicago orthologues with expression data	***Arabidopsis* TF**	**Chromosome no. and co-ordinates**	**Medicago mutant line identified**
Medtr7g118330	CCA1	chr7:49115428-49124034	NF2784, NF17115, NF20691, NF16461
Medtr5g076960	RVE1	chr5:32828190-32832068	NF13921
Medtr2g094520	ATHB15	chr2:40311598-40317520	NF1297
Medtr3g109800	ATHB15	chr3:51353722-51359959	NF5993, NF17550
Medtr2g030130	ATHB15	chr2:11282450-11288303	NF0908, NF1084, NF1729
Medtr3g109800	ATHB15	chr3:51353722-51359959	NF14210
Medtr8g013980	ATHB15	chr8:4335616-4343529	NF11854
Medtr1g079810	AHL20	chr1:35464260-35466532	NF1531, NF17813
Medtr4g098450	AHL20	chr4:40540545-40542860	NF2932, NF20547
Medtr5g011520	AHL20	chr5:3331246-3333409	NF9265, NF3697
Medtr3g100470	AHL20	chr3:46200293-46202285	NF17277
Medtr5g091630	AHL20	chr5:39959279-39960443	NF11111, NF1264_2D
Medtr7g080980	AHL25	chr7:30873014-30875062	NF3772, NF5835, NF13955, NF20437

Table 5.1.2-2: Tnt1 mutant lines identified from the blastn search from the Medicago mutant database (Noble Research Foundation). Medicago genes corresponding to orthologous *Arabidopsis* TF(s), FST start and end site co-ordinates with the respective mutant lines found for each gene are given below. The list of orthologous genes in this table were not found to be expressed in our microarray data.

Medicago orthologues without expression data	*Arabidopsis* TF	Chromosome no. and co-ordinates	Medicago mutant line identified
Medtr3g109800.2	ATHB15	chr3:51353736-51359959	NF5993
Medtr4g058970.1	ATHB15	chr4:21738748-21744420	NF5999
Medtr8g089895.1	ATHB16	chr8:37518358-37521844	NF14591
Medtr1g073860.1	AHL20/ AHL25	chr1:32802741-32804858	NF8126, NF12227
Medtr1g044155.1	AHL20/ AHL25	chr1:16576895-16579794	NF19358,
Medtr3g068035.1	AHL20/ AHL25	chr8:13212033-13213496	NF17814, NF17813
Medtr8g036060.1	AHL20/ AHL25	chr3:54113534-54116609	NF7335 NF10935
Medtr3g115620.1	WOX13	chr3:54113534-54116609	NF19167, NF21293

5.1.3 Genotypic screening of *Tnt1* inserted mutant lines identify homozygous and wild-type sibling plants

Around 15-20 seeds for each of the 31 mutant lines were purchased from the Noble Research Foundation, US. All of these Medicago mutant lines were grown in the glasshouse after germination. On average 10-15 plants for each line germinated and had wild-type level growth in glasshouse conditions. In total, around 500 Medicago plants were genotyped to screen for the presence of FST amplicons. PCR-based screening with pairs of gene-specific primers (GSP) and *Tnt1* primers were used to identify the homozygous and wild-type sibling plants for each of the individual lines. Four primer combinations for each orthologous genes: GSP-F with Tnt1-F, GSP-F with Tnt1-R, GSP-R with Tnt1-F, and GSP-R with Tnt1-R was used for screening by following touch-down PCR program.

All the working gene specific primer sequences used for genotyping and PCR conditions are given in Method section 2.5, table 2.5-1 and 2.5-2.

Genotyping for the presence of *Tnt1* inserts and confirmation of the homozygosity with PCR has been presented for eight mutant lines in this section: NF17115 and NF16461 for *Medtr7g11830*, NF13921 for *Medtr5g076960*, NF5993 for *Medtr3g109800*, NF1729 and NF0908 for *Medtr2g030130*, NF5999 for *Medtr4g058970* and NF12227 for *Medtr1g073860*.

Out of the six lines with *Tnt1* inserts detected from a pool of 12 plants in NF17115, two homozygous lines were identified (Figure 5.1.3-1, B1 & B2). Mutant lines with inserts showed the expected size amplification (~1 kb) when PCR was performed with the *Tnt1*-specific and GSP primer pair. Out of these six lines with inserts, plants 2 and 11 were confirmed as homozygous plants as a WT band could not be amplified. Plants 1, 4, 5 and 9 were confirmed as heterozygous because they showed similar amplification with expected product size of ~1.5 kb as in the wild-type R108 control (Figure 5.1.3-1, B2). This approach of PCR screening and confirmation was used as the method for genotyping all the remaining mutant lines.

From the eight mutant lines presented here, three homozygous plants were identified for NF16461 (plant no. 157, 158, 166, Figure 5.1.3-1, C1 & C2); one homozygous line for NF13921 (plant no. 25, Figure 5.1.3-2, A1 & A2); four homozygous lines for NF5993 (plant no. 95, 100, 107 and 111, Figure 5.1.3-2, B1 & B2); two homozygous lines for NF1729 (plant no. 170 and 175, Figure 5.1.3-3, B1 & B2); two homozygous lines for NF1729 (plant no. 189 and 190, Figure 5.1.3-3, C1 & C2); two homozygous lines for NF5999 (plant no. 120 and 121, Figure 5.1.4-4, A1 & A2). Wild-type siblings (WTS) were also isolated from these lines.

Homozygotes could not be detected in some of the mutant lines screened and progenies were found to be all heterozygous with few wild-type sibling lines. One example for such a case is presented here for the mutant line NF12227. Out of the 7 Tnt1-insert positive plants (plants no. 45, 49, 50, 52, 55, 56, 61) from 20 plants, all of them were heterozygous lines.

From seven lines Tnt1 inserts could not be amplified, thus these were identified as WT because they only amplified a control R108/WT sized product (~2 kb) (Figure 5.1.4-4, B1 & B2).

16 mutant lines representing eight Medicago ortholog genes of the *ATHB* and *AHL* gene family were not genotyped due to time constraints (Figure 5.1.3-5).

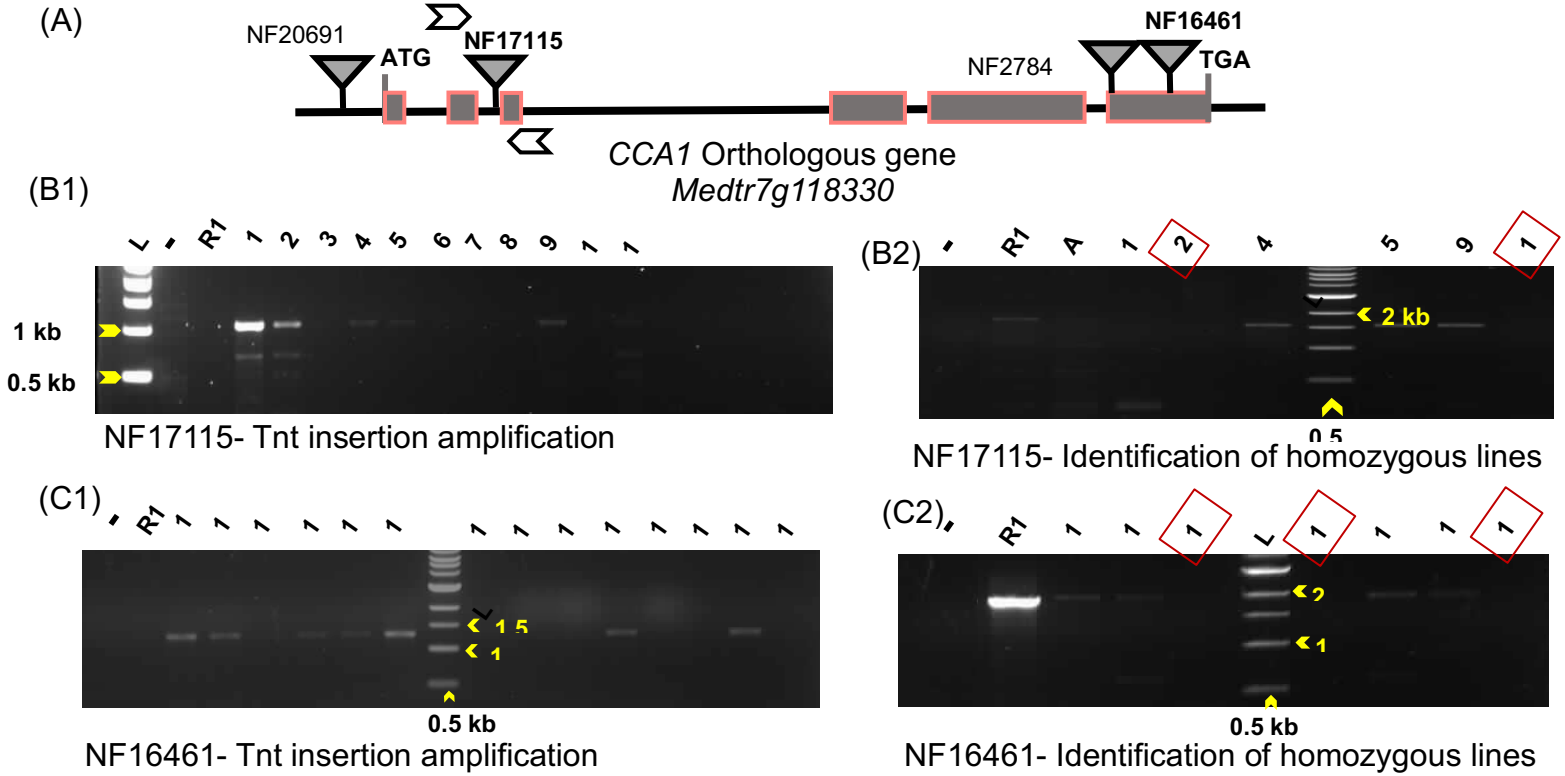


Figure 5.1.3-1 Molecular characterisation of mutants in Medicago CCA1 orthologous gene *Medtr7g118330*. (A) Four insertion sites in the *Medtr7g118330* genomic sequence. Filled grey boxes with pink borders represent exons while the solid black line represents introns. Triangle positions denote *Tnt1* insertion sites, as identified from the nBLAST search in the mutant database. Arrow positions indicate the site of primer design for qPCR-based quantification. NF stands for Noble Foundation where *Tnt1* lines were generated in the WT-R108 background**. (B1)** PCR amplification for Tnt inserts with Tnt primer TntFg and CCA1-2-Rp1 in NF17115. Plants 1,2,4,5,9 and 11 harbour the NF17115 *Tnt1* insertion. **(B2)** PCR screen for homozygosity with GSP pair CCA1-2-Fp2 and CCA1-2-Rp1 showing plant 2 and 11 as homozygous lines with no detection of amplification of the WT band. **(C1)** PCR amplification for Tnt inserts with Tnt primer TntFg and NF16461-Fp1. Plants 154, 155, 157, 158, 159, 163 and 167 harbour the *Tnt1* insertion. **(C2)** Plant 157, 158 and 166 were identified to be homozygous lines by testing for the absence of a WT band using GSP pair NF16461-Fp1 and NF16461-Rp1. Plant 161 was found to be a WTS (not shown); see Methods, Table 2.5-2 for primer sequences.

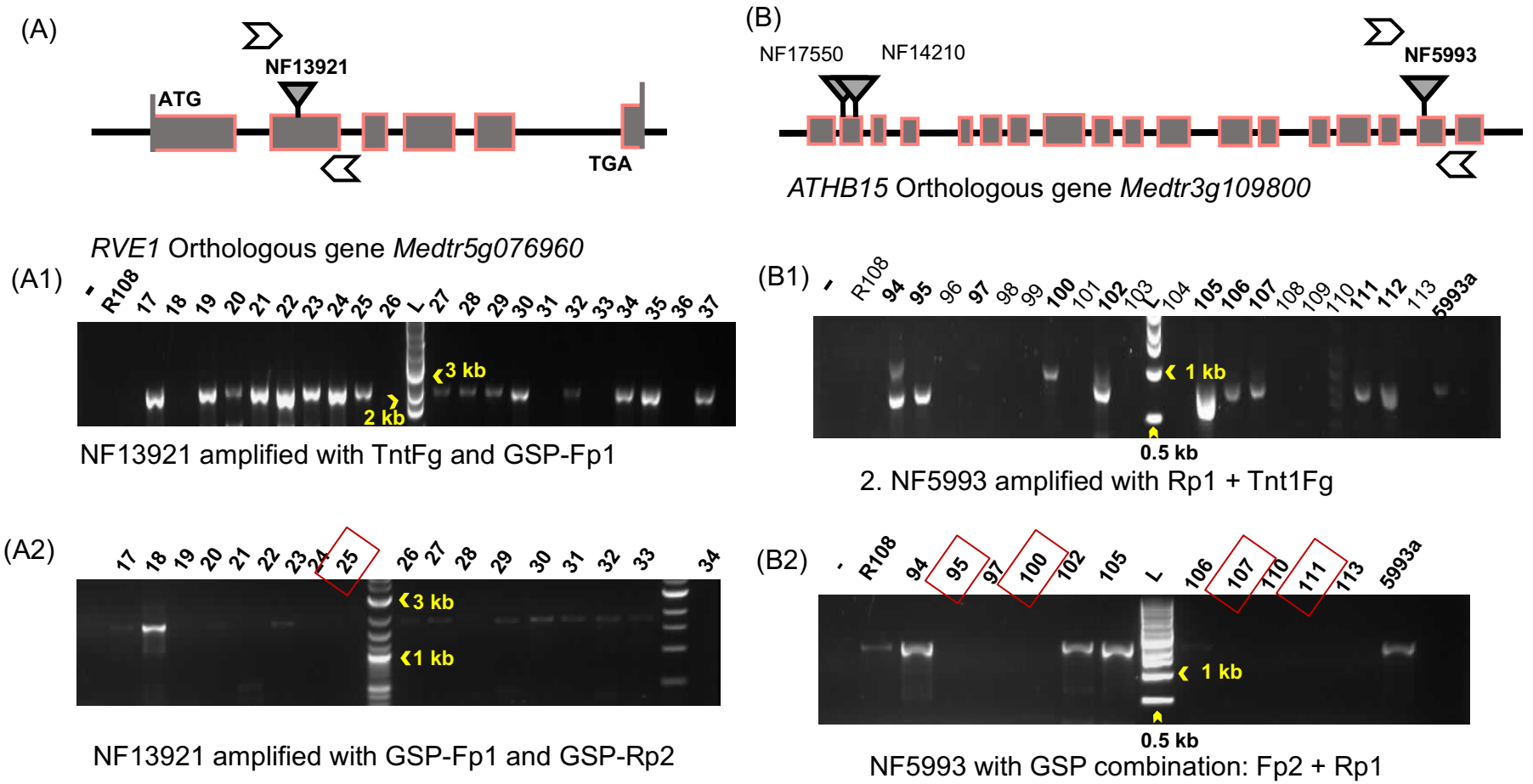


Figure 5.1.3-2 Molecular characterization of RVE1 Medicago ortholog gene *Medtr5g076960* and ATHB15 ortholog *Metr3g109800*. (A) One *Tnt-1* insertion (denoted by triangle position) in the *Medtr5g076960* genomic sequence. Arrow positions indicate the site of primer design for qPCR-based quantification. (B) Three Tnt1-1 insertions in Medtr3g109800. Homozygous lines (highlighted with red boxes) were selected for further analysis. (A1) PCR amplification for Tnt inserts with Tnt primer TntFg and GSP-Fp1 in NF13921. (A2) Plants 25 putatively confirmed as homozygous line and 18 as WTS. (B1) PCR screen for Tnt inserts in NF5993 (B2) Homozygosity confirmation in plant line no. 95, 100, 107 and 111 OF NF5993 with GSPs pair Fp2 and Rp1.

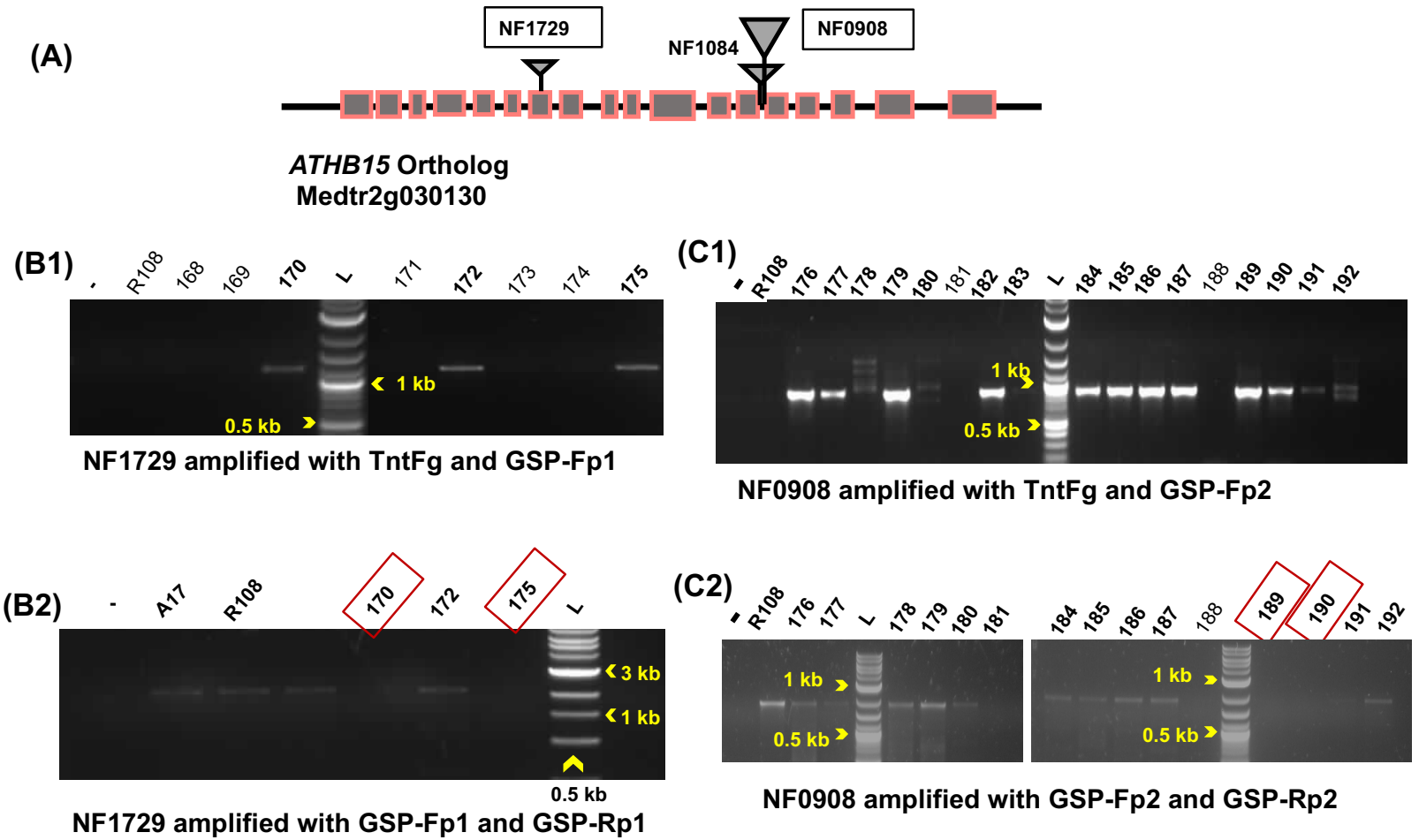


Figure 5.1.3-3 Molecular characterization of ATHB15 Medicago ortholog gene *Medtr2g030130*. (A) Three *Tnt-1* insertions (denoted by triangle position) in the *Medtr2g030130* (B1) Tnt insertions confirmation in NF1729 with TntFg and GSP-Fp1 (B2) Homozygosity confirmation in line 170 and 175 with GSP-Fp1 and GSP-Rp1 in NF13921. (C) Tnt insert confirmation in NF0908 with TntFg and GSP-Fp2. (C2) Homozygosity confirmation in plant line no. 189 and 190 with GSPs pair Fp2 and GSP-Rp2

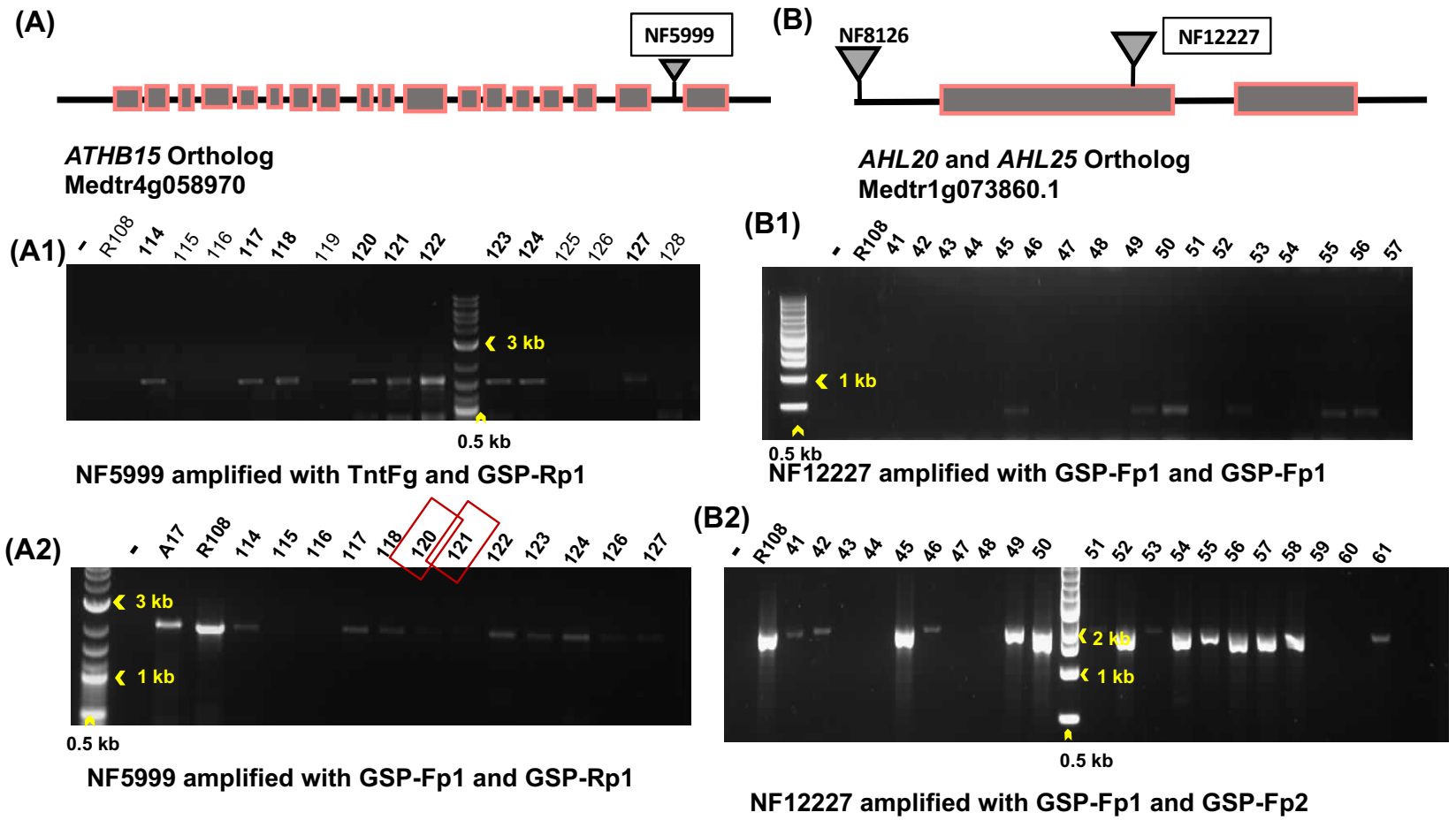


Figure 5.1.3-4 Molecular characterization of ATHB15 Medicago ortholog gene *Medtr4g058970*. (A) One *Tnt-1* insertion in the *Medtr4g058970*. (A1) Tnt insertions confirmation in NF5999 with TntFg and GSP-Rp1 (A2) Homozygosity confirmation in line 120 and 121 of NF5999. (B) Tnt insertion in *Medtr1g073860*. (B1) Tnt confirmation with TntFg and GSP-Fp1 (B2) All heterozygous lines in NF12227 confirmed with GSP-Fp1 and GSP-Fp2.

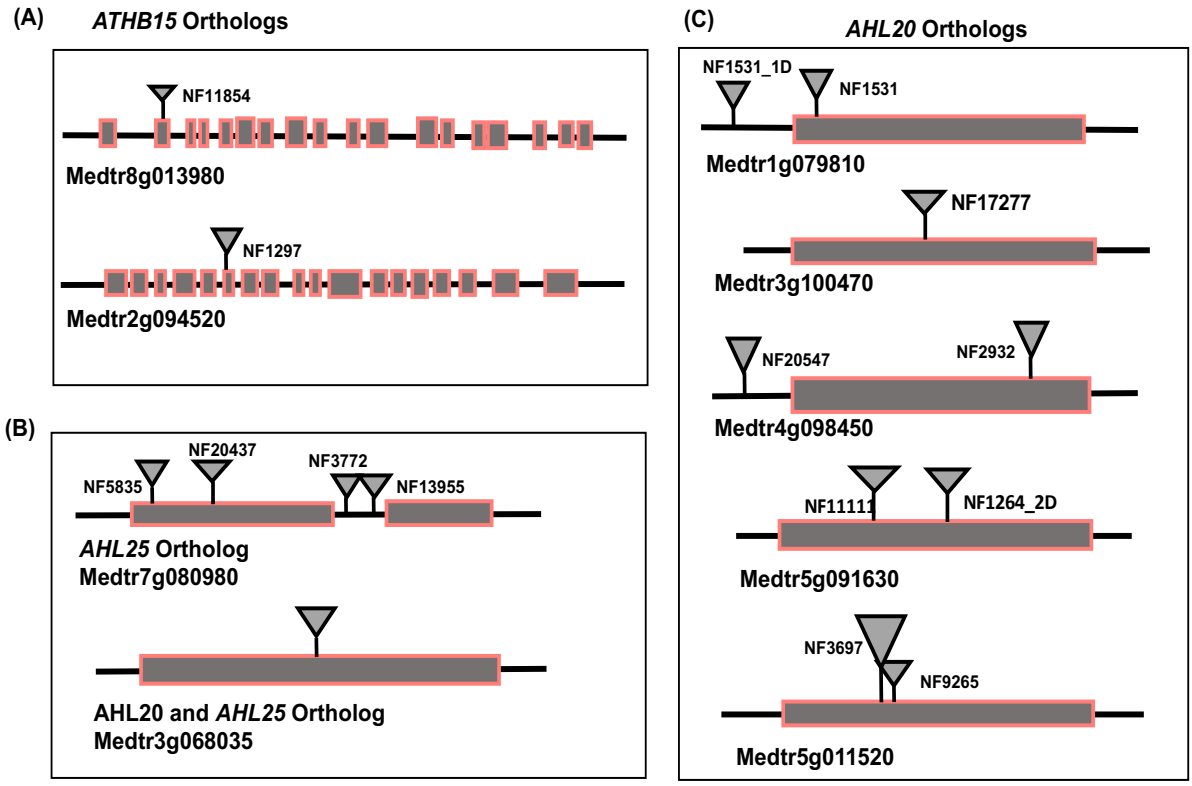


Figure 5.1.3-5 ATHB15, AHL20, AHL25 Medicago ortholog genes with their Tnt insertion(s) in their genes. Filled grey boxes with pink borders represent exonic region while solid black lines represent intronic regions. The position of the triangle denotes the site of Tnt insertions in the gene as identified with nucleotide BLAST search from the Medicago mutant database. NF stands for Noble Foundation where these Tnt lines were generated in the wild type R108 background. These lines were not analysed for PCR screening and qPCR quantification.

5.1.4 Relative quantification of mRNA expression in wild-type sibling and homozygous mutants by qPCR

After successful confirmation of *Tnt-1* inserts presence and homozygosity in 7 *M. truncatula* mutant lines (NF17115, NF16461, NF13921 for *CCA1* or *RVE1* ortholog genes; NF5993, NF1729, NF0908 and NF5999 for *ATHB15* ortholog genes) by PCR screening, these plants were grown to maturity in the glasshouse and seeds were collected. If any WTS plant lines were identified from the genotyping, their seeds were also collected.

Seeds of these homozygous lines, WTS-lines and wild-type R108 as control were germinated and grown on Fahraeus media plates (see Methods). *athb15* mutant line seedlings were grown for 7 days and *cca1* and *rve1* mutant line seedlings for 25 days due to slower growth of these lines. Root samples were then harvested and used for mRNA extraction and cDNA synthesis. SYBR green dye based quantitative PCR (qPCR) was used to quantify TF gene mRNA expression in the transcribed cDNA from the root samples of homozygous, wild-type siblings, and R108 (control) Medicago seedlings.

The expression profile of homozygous lines representing the four Medicago orthologous genes *Medtr7g11830*, *Medtr5g076960*, *Medtr3g109800* and *Medtr4g058970* were investigated with these primers. Target specific qPCR primers designed on the exonic region of the orthologous genes were tested for efficiency using the calibration curve method (Table 5.1.4-1) and melt curve analysis (Figure 5.1.4-1). Two reference genes (*β-Tubulin* and *MtPDF2*) were used as control genes for normalisation (Kakar et al., 2008); all primers sequences are listed in Methods section 2.5-1 and 2.5-2.

For the *Medtr7g11830 CCA1* ortholog gene fold change analysis in the mRNA expression between WT-R108 and two homozygous mutant lines (plant no 11 and 2) for NF17115 confirmed loss of mRNA expression in all of the homozygous lines. Expression of the *Medtr7g11830* mRNA was reduced by 3.5 fold and 7.5 fold in homozygous plant 11 and 2 respectively as compared to mRNA from WT-R108 root samples (Figure 5.1.4-2 A).

In a second (allelic) mutant line (NF16461) for the same *CCA1* ortholog gene, *Medtr7g11830*, there was also lower *Medtr7g11830* mRNA expression in homozygous plant no 157 and 158 (Figure 5.1.4-2 B). One of the

homozygous line (Ho158) displayed a FC reduction of 6.6 as compared to WT-R108. However, the WTS-161 for mutant line NF16461 (Figure 5.1.4-2 B) did not display a similar expression profile to that of wild-type R108 plant, and thus this result must be treated with some caution. A possible explanation of this could be due to the location of the primer design. The same qPCR primer pair designed for NF17115 line was used for mRNA quantification of NF16461. NF16461 insert lies on the 6th exon while NF17115 insert is in between 2nd and 3rd exons of *Medtr7g11830*. Hence, use of these primers for quantification of mRNA in the WTS line and homozygous 157 plant for NF16461 (Figure 5.1.4-2 B) might not have amplified the most informative region.

For the NF13921 line carrying a Tnt1 insertion in the *RVE1* ortholog *Medtr5g076960*, homozygous plant no. 25 had a strong reduction of *Medtr5g076960* mRNA expression level as compared to the WT-R108 plant (Figure 5.1.4-2 C). The WTS line for this mutant line also displayed reduced mRNA expression as compared to WT-R108 (Figure 5.1.4-2 C), thus further analysis is of this line is necessary.

For *Medtr3g109800* gene, representing one *ATHB15* ortholog gene, two of the homozygous lines for NF5993 showed significant reduction of *Medtr3g109800* mRNA expression level (Figure 5.1.4-3 A) compared to WT-R108. A homozygous line for the *ATHB15* orthologous gene *Medtr4g058970* (NF5999) was found not to have a significant altered *Medtr4g058970* mRNA expression level (Figure 5.1.4-3 B) suggesting that the presence of the Tnt1 insertion does not disrupt gene expression. However, the primer pair designed to quantify mRNA of *Medtr4g058970* was subsequently found not to span the region of the Tnt insert in NF5999 line, thus use of these for the homozygous 115 and WTS plant lines of NF5999 is possibly unreliable. New primer pairs for NF16461 and NF5999 will be designed and optimised as part of future work to perform complete validation of these two lines.

Table 5.1.4-1: Five Medicago ortholog genes analysed for perturbed mRNA expression by quantitative real-time PCR (qPCR). Table shows the efficiency of primers designed from the calibration-curve method (data plot not shown) for each mutant line; primer efficiency ranges from 1.7-2.1. Two reference house-keeping genes (*β-Tubulin* and *MtPDF2*) used for normalisation as controls are also listed. Mutant lines in bold were fully analysed using qPCR. Chr2/3/4 denotes the chromosome number on which the ortholog genes are located.

***Medicago truncatula* genes**	**Slope**	**log E (1/slope)**	**Primer efficiency (E)**	**Target Medicago mutant lines**
β-Tubulin *Medtr7g089120* (Kakar et al., 2008)	3.6376	0.27490653	1.88324374	-
MtPDF2 *Medtr6g084690* (Picard et al., 2013)	3.4581	0.28917614	1.94614922	-
CCA1 *Medtr7g118330*	3.6556	0.27355291	1.8773831	**NF17115**, **NF16461**
RVE1 *Medtr5g076960*	3.4907	0.28647549	1.93408471	
ATHB15 on chr2 *Medtr2g030130*	4.2554	0.23499554	1.71789073	**NF13921**
ATHB15 on chr3 *Medtr3g109800*	3.6916	0.27088525	1.86588663	NF1729, NF0908, NF1297
ATHB15 on chr4 *Medtr4g058970*	3.0623	0.32655194	2.12105504	**NF5993**, **NF5999**

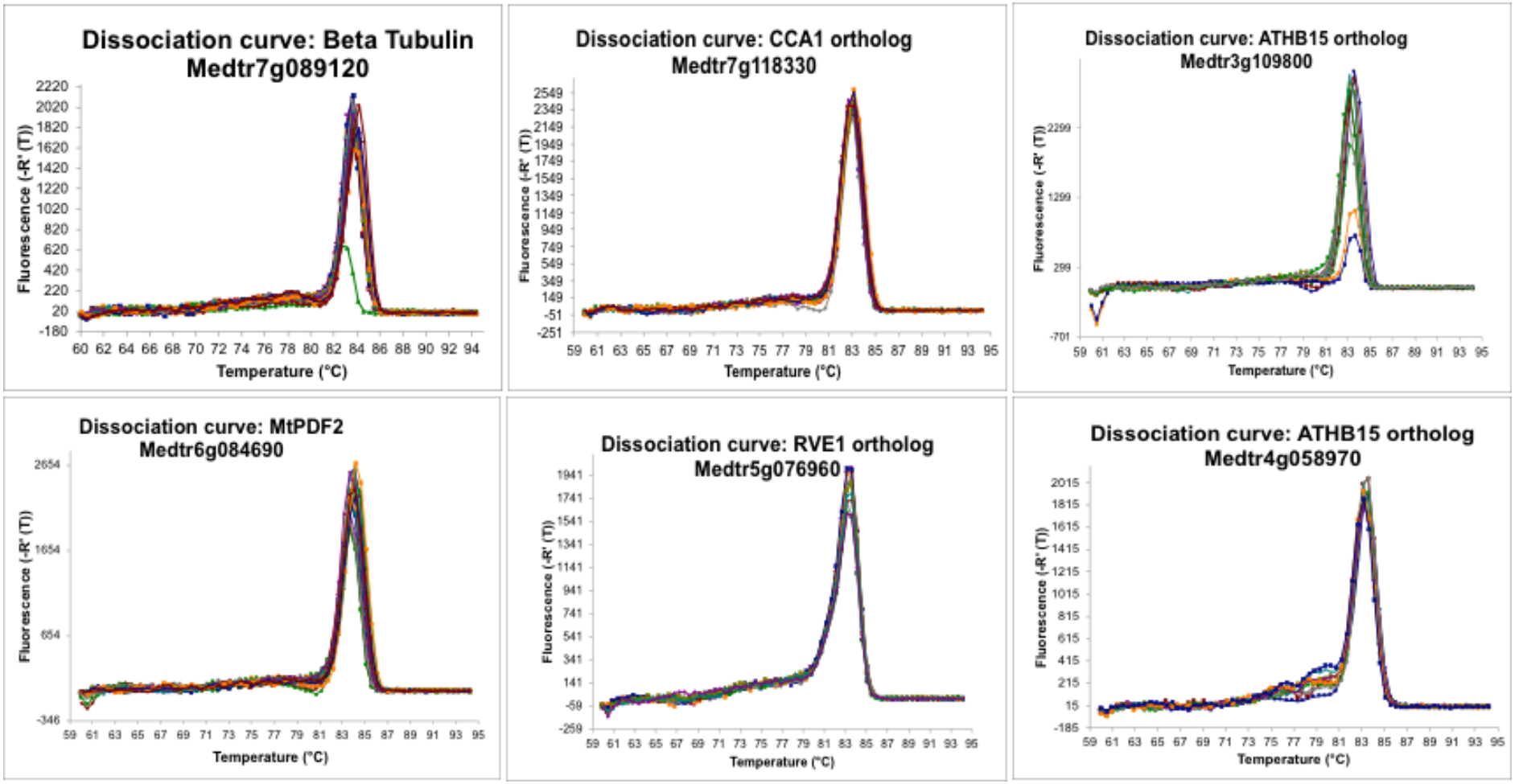


Figure 5.1.4-1 Melt curve analysis from the real time qPCR assay for the six primer pairs used for mRNA quantification. Primers used (see Methods) for amplification reactions for Medtr7g118330 (NF17115-Ho11, Ho2; NF16461-WTS161, Ho157, Ho158); Medtr5g076960 (NF13921- WTS18, Ho25); Medtr3g109800 (Ho111, Ho107, Ho95); Medtr4g058970 (Ho115, WTS116); coloured lines denote individual reactions.

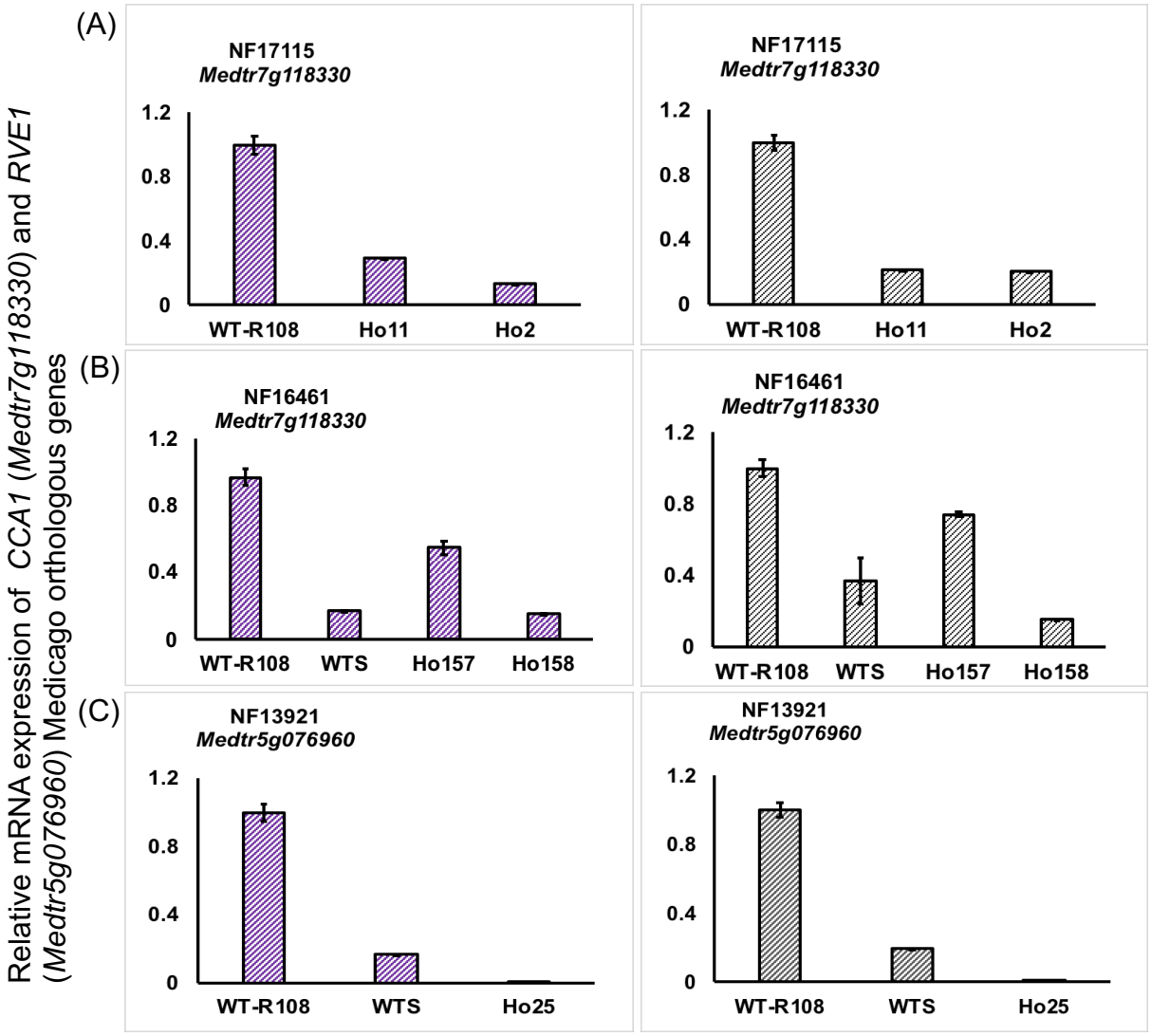


Figure 5.1.4-2: qPCR analysis of *CCA1* ortholog *Medtr7g11830* and *RVE1* orthologous gene *Medtr5g076960*. Gene expression in wild-type R108, homozygous and WTS lines identified from genotyping were normalised using two reference genes, (1) *β-Tubulin Medtr7g089120* and (2) *MtPDF2 Medtr6g084690*. WT- wild type, Ho- homozygous, WTS- wild type sibling. Data represent the geometric mean of ΔCt, normalised and calculated as described in Methods. Error bars represent standard deviations of FC of the mean ΔCt values (n=3). **(A)** Relative mRNA expression of the *Medtr7g11830* gene in the mutant lines of NF17115 (WT-R108, homozygous- Ho11 and Ho2). No WTS was detected in NF17115 lines from the genotypic screen. (B) NF16461 (WT-R108, WT sibling, homozygous lines- Ho157 and Ho158) **(C)** Relative mRNA expression of *Medtr5g076960* gene in the mutant lines of NF13921 (WT-R108, WTS-18, Ho25). One WTS each was detected for NF16461 and NF13921 lines.

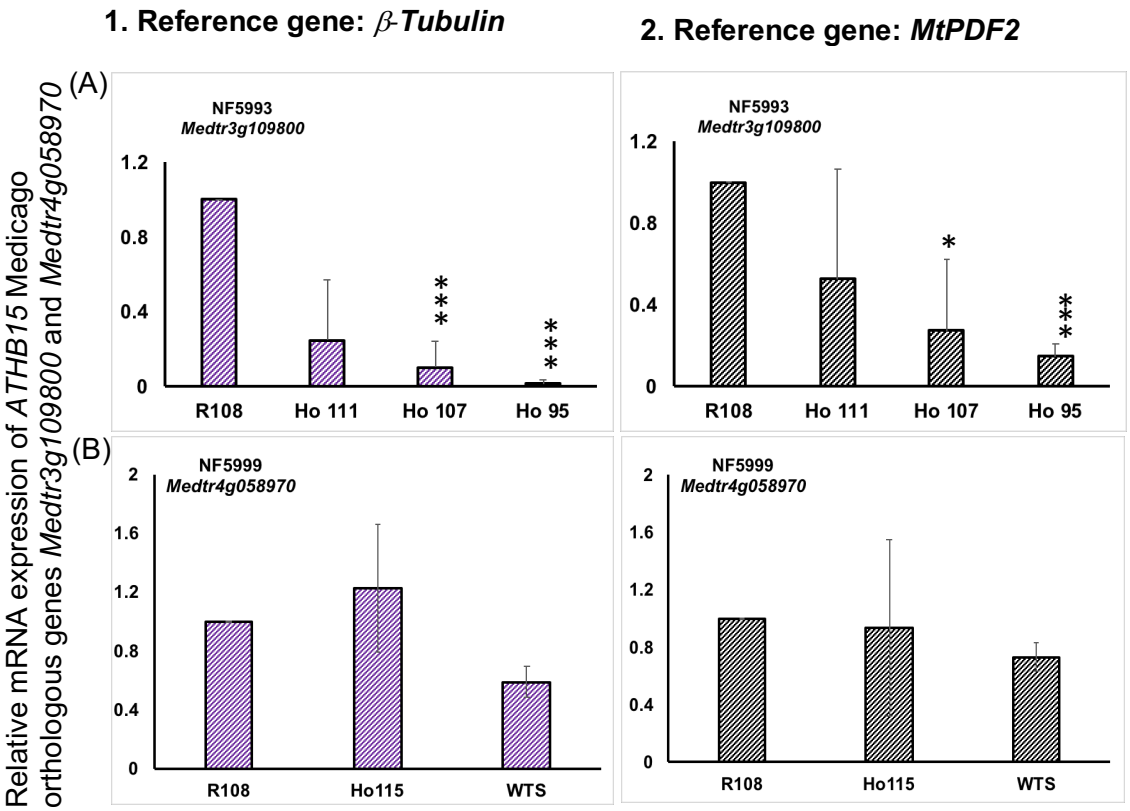


Figure 5.1.4-3: qPCR analysis of *ATHB15* orthologous genes *Medtr3g109800* and *Medtr4g058970*. Gene expression in wild-type R108, homozygous and WTS lines identified from genotyping were normalised using two reference genes, (1) *β-Tubulin, Medtr7g089120* and (2) *MtPDF2, Medtr6g084690*. WT- wild type, Ho- homozygous, WTS- wild type sibling. Data represent geometric mean of ΔCt, normalised and calculated as described in Methods. Error bars represent standard deviations of FC of the mean ΔCt values (n=2) **(A)** Relative mRNA expression of *Medtr3g109800* gene in the mutant lines of NF5993 (WT-R108, homozygous- Ho111, Ho107 and Ho95). No WTS line was detected from the screen of NF5993 mutant population. **(B)** Relative mRNA expression of *Medtr4g058970* gene in the mutant line NF5999 (WT-R108, Ho115, WTS). One WTS line was detected from the genotypic screen.

5.2 Discussion

In Chapter 5, the Medicago mutant database developed at the Noble Research Foundation, USA was used for the search of *Medicago truncatula* mutants. The *Medicago truncatula* genome is four times larger than the *Arabidopsis* genome and this is the only easily accessible large-scale efficient T-DNA transformant collection (Somers et al., 2003; Tadege et al., 2008). The Tnt1 mutant population database offered the opportunity for gene characterisation and legume functional genomics.

A BLAST search of the population of high-confidence Tnt1 mutants was used based on the flanking sequences of the Tnt-1 insertions, to identify Medicago mutant lines disrupted in the orthologous genes. PCR screening in the progenies of the mutant lines obtained for CCA1, RVE1, ATHB15 orthologs successfully confirmed the presence of Tnt1 inserts using the Tnt1-specific primers. The expression of perturbation of circadian associated Medicago orthologous genes *Medtr7g118330* (CCA1) and *Medtr5g076960* (RVE1) was confirmed to be perturbed in the homozygous lines.

Progenies from these homozygous plants can be used for further investigation to study the involvement of the circadian system in NCR expression during Medicago-rhizobium symbiosis. For example, the homozygous lines of NF17115 and NF16461 (for *Medtr7g11830*) and NF13921 (for *Medtr5g076960*) are candidates for further investigation to examine the effect of rhizobia and N treatment on NCR expression when circadian TF(s) are perturbed. The mutant line NF5993 of the ATHB15 ortholog *Medtr3g109800*, currently annotated as a class III homeodomain leucine zipper protein, is another candidate to study since there is a significant loss of *Medtr3g109800* expression in the two homozygous plant lines Ho95 and Ho107.

A recent ATAC-seq study for comparison of chromatin root profiles in four plant species including *M. truncatula* identified ‘expressologs’ (functional homologs) for important genes known to be involved in root development (Maher et al., 2018). Their findings implicated possible relocation of the cis-regulatory elements in the upstream region over evolutionary time but stabilisation and conservation of TF(s) regulatory control across species. Their data include *Medtr3g109800* as an expressolog of root developmental genes.

This supports the prioritization of analysing this gene in further work and provides a hypothesis, interaction as TF(s) with its binding sites in the NCR gene family, for the role of *Medtr3g109800* in root developmental responses.

Overall the molecular characterization of the homozygous mutant lines presented in this chapter and confirmation of reduced mRNA expression by qPCR in many of them has enabled us to start to analyse the loss of function in our TFs of interest. Although phenotypic data is not presented here, preliminary phenotypic observation (data not shown) of these homozygous mutant lines showed different developmental root response timing. *cca1* and *rve1* ortholog homozygous mutant lines analysed here (NF17115, NF16461 and NF13921) needed to be grown for a longer time until there was proper root development as compared to WT-R108 and the *ATHB15* mutant lines. While 7 day old seedlings of R108, A17 and *athb15* mutants NF5993 and NF5999 showed proper root development (in terms of PR & LR emergence), *cca1* and *rve1* homozygous mutant lines analysed in our condition showed reduction in growth rate. This was the reason for qPCR mRNA quantification at different growth stage for the circadian associated mutants and *athb15* mutants. It will be very interesting to characterise these mutants in further work to ask if there is a root development related phenotype.

In this work we have demonstrated the utility of the Tnt Medicago mutant population for forward genetics by identifying insertion mutations in the orthologous genes of our putative TFs. The circadian and plant development related orthologous genes that were identified need to be functionally characterized using phenotypic analysis. As part of this the mutant lines should be back-crossed to the parental R108 in order to develop single mutant copy lines. Cosegregation analysis to remove the extraneous mutations should also be performed (as carried out in (Weigel and Glazebrook, 2008)). Complementation (using cotyledons for somatic organogenesis and thus rapid transformation) may also be useful in cases where we fail to find the alleles by evaluating the candidate FSTs of the mutant phenotype (Tadege et al., 2008; Zhou et al., 2004). In conclusion, the molecular characterisation of such Medicago Tnt mutant lines created an extremely useful resource.

Chapter 6: Time course analysis of gene expression to ask if NCR genes oscillate in a circadian manner

6.1 Introduction

In chapters 4 and 5, we found that NCR gene clusters had an overrepresentation of AGAT/CATTT elements that were target sites for circadian TF binding. We had also presented the outcome of a search of *Medicago truncatula* mutant lines generated in the orthologous genes *CCA1, RVE1* and *ATHB15* to be used in molecular characterization. To ask if the NCRs could be under circadian control we sought to profile the transcriptome of the wild-type A17 Medicago over a 48 hour time period and analyse NCR gene family expression.

Biological circadian clocks are internal time-keepers that have been evolved in all forms of organism. Among plant species, most of present knowledge of the circadian clock comes from the widely studied non-legume model *Arabidopsis thaliana* (Song et al., 2010) and limited reports in the non-IRLC legume *Glycine max* (Chiasson et al., 2014; Marcolino-Gomes et al., 2014). Conservation of other photoperiodic genes such as *EARLY FLOWERING* (*ELF3*, *ELF4* and *ELF6*) and *GIGANTEA*, *GI* along with the representation of *CCA1* and *LHY* by single sequence have been reported in soybean and *Medicago truncatula* based on an EST comparative (Hecht et al., 2005).

Internal biological circadian clocks drive molecular rhythms temporally with anticipatory changes in the environment. Till this date, there has not been a comprehensive study of circadian control in N-fixing nodules (Chiasson et al., 2014). In this context, we propose that with the conserved CCA1 and RVE1 binding site present in the NCR promoters, a nodule circadian component from the host plant could be involved to enable coordinated regulation of rhizobial colonisation and responses external N availability.

We hypothesised that circadian control in a nodule will be similar to that of roots (James et al., 2008) but under the influence of N and rhizobia. To test our hypothesis for a ‘nodule circadian clock’, we generated a transcriptomic

time-series data of nodulating *Medicago truncatula* roots to study the regulation of NCR expression under the combinatorial effect of nitrogen status and rhizobia. A preliminary analysis of the transcriptome is presented here.

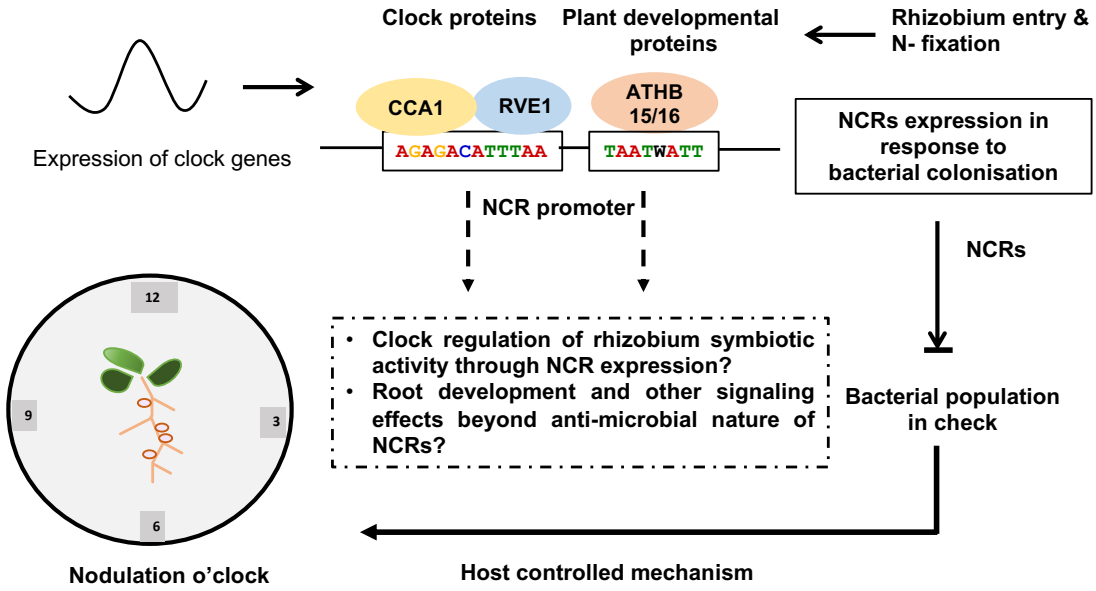


Figure 6.1: A proposed model for the regulation of the NCR gene expression by circadian clock and plant development proteins. As rhizobial entry at low N conditions leads to nodulation, *NCR* gene expression could be under transcription factor control to maintain a symbiotic rhizobium balance in host nodules. Black dotted arrows represent the downstream activity that can influence the level of rhizobium symbiosis and subsequent root architecture development under the host controlled and tight regulation of these TF(s)-NCR promoters.

6.1.2 Expression pattern of NCR genes with degenerate circadian CCA1 or RVE1 motif binding sites

To evaluate the existence of circadian regulation in normal day-night cycling growth conditions, nodulating seedlings (4 day old rhizobia-inoculated) of *Medicago truncatula* wild-type A17 were grown at 16 hours day and 8 hours light cycles (Figure 6.1.2-1). We chose a time point of 7 dpi for nitrogen treatment at deplete (0.1 mM) and replete N (5 mM) concentrations because previous findings reported that full activation of the NCR transcriptome takes place at 6 dpi to 10 dpi (Guefrachi et al., 2014).

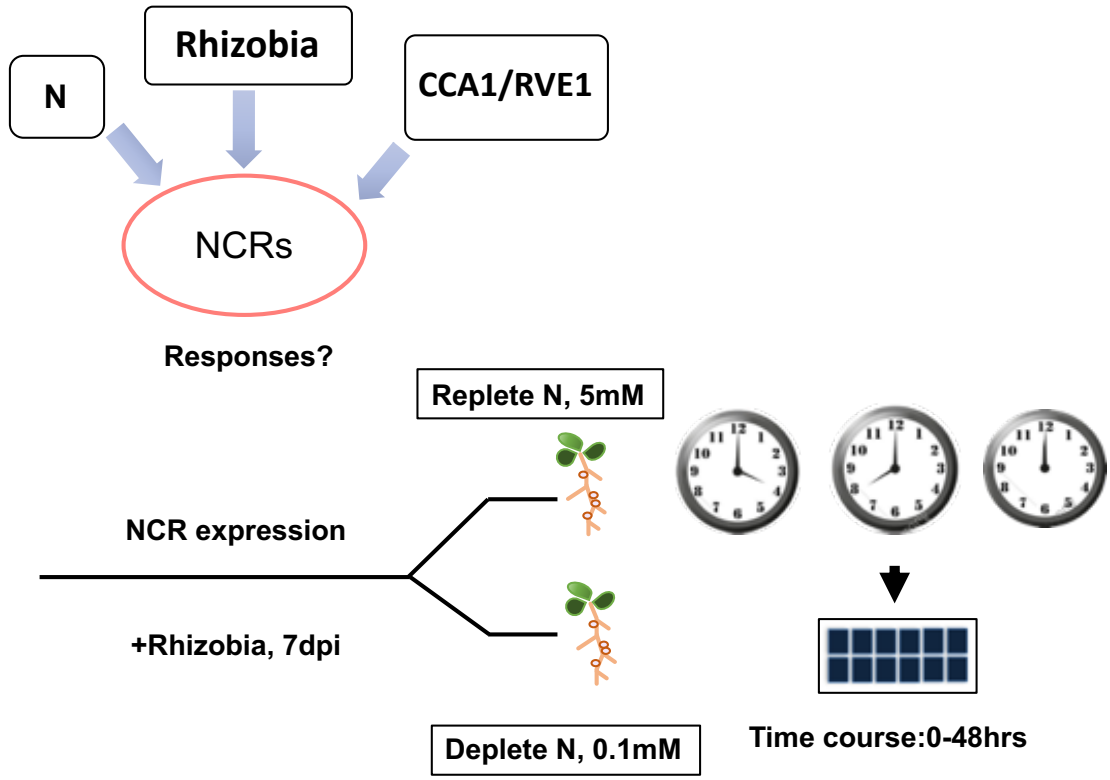


Figure 6.1.2-1: Overview of the experimental approach for the Medicago wild type A17 plant time-series from 0-48hours. Medicago A17 plants were grown in deplete N conditions. *Sinorhizobium meliloti* Sm1021 was then inoculated on 4 d old seedlings. 7 dpi plants were then transferred to deplete (0.1 mM) nitrogen and replete (5 mM) nitrogen conditions. The sampling interval of roots for RNA extraction was 4 hours for a period of 0-48 hours in both deplete and replete nitrogen conditions. All the plants were grown at 16 hours light and 8 hours dark in 25 °C conditions. For the root architecture studies (data not shown) plant growth was continued in deplete or replete N condition for 15 more days.

Nodulating roots were sampled for mRNA extraction with an interval of every 4 hours until 48 hours for both deplete and replete N conditions. The experimental approach for the time-series transcriptomic analyses is given in Figure 6.1.2-1. Samples were analysed using microarray hybridization, with data normalized (see Methods).

Our preliminary results from the transcriptomic analyses focused only on the expression of the 185 NCRs found to be expressed in previous chapters. Before examining their expression, the promoter region of all the 185 NCRs [-1,-500] was scanned for the presence of circadian CCA1 or RVE1 motif sites using FIMO (see Methods). There were 92 NCRs that contained a single CCA1 or RVE1 motifs with degenerate binding sites, and 41 NCRs that contained multiple motif sites. 52 NCRs were found with no motif sites of CCA1 or RVE1 TF(s), however these NCRs contained other motif combinations that we described in Chapter 4. The expression pattern of the 133 NCRs that had single or multiple motif binding sites for CCA1 and RVE1 was then plotted over time (Figure 6.1.2-2).

All NCRs, irrespective of whether a single or multiple motif sites were present, showed variation in the gene expression pattern over a period of 48 hours in both deplete and replete N conditions (Figure 6.1.2-2). NCRs in deplete N conditions demonstrated a wide variation of expression levels but all appeared to have some oscillatory variation (Figure 6.1.2-2, C and D). In N replete conditions, NCRs exhibited reduced variation or damping in the expression level towards the end of the 48 hour period, relative to their expression at 0 hour (Figure 6.1.2-2, A and B). Such an expression pattern was not overserved in the N deplete conditions.

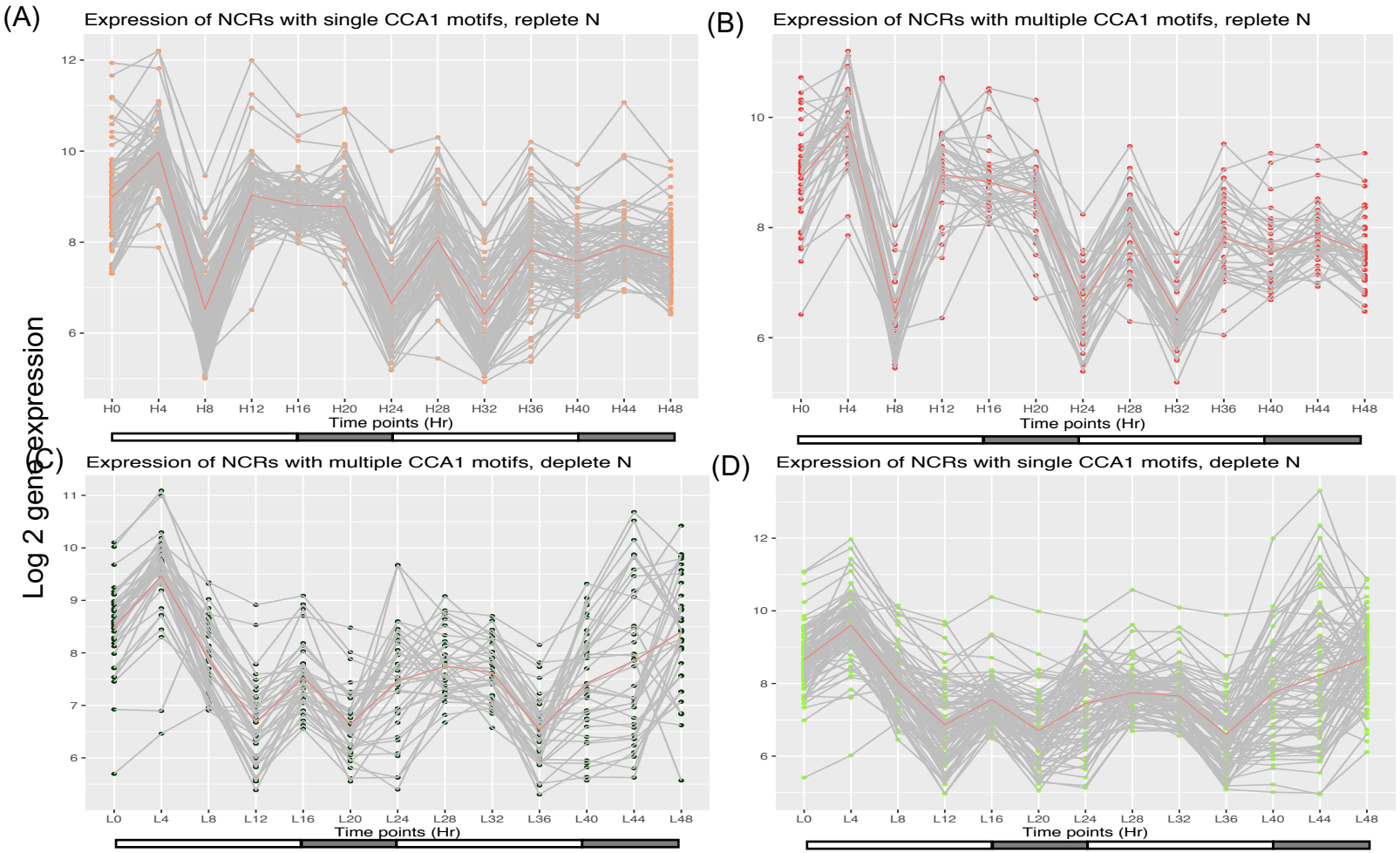


Figure 6.1.2-2: **Expression pattern of NCR promoter genes from the A17 time series transcriptomic data.** Unfilled rectangle box below X-axes represent 16 hours of light and the grey solid box represent 8 hours of dark. Red curve in the middle represents the average gene expression of the *NCR* genes. Time points in hours; L0-L48 at deplete N (0.1 mM) conditions; H0-H48 time points at replete N (5 mM). CCA1-TFBS: CCA1 transcription factor binding sites. **(A)** NCR promoters with single motif sites, replete N. **(B)** NCR promoters with multiple motif sites, replete N. **(C)** NCR promoters with single motif sites, deplete N. **(B)** NCR promoters with multiple motif sites, deplete N.

6.1.3 Expression pattern of *NCR* genes with core AGATATTT element

To further investigate any expression pattern due to interaction of nitrogen, CCA1 and RVE1 with the NCR promoters that contained the core and exact AGATATTT binding element, we sought to evaluate the expression of 15 *NCRs* that contained this element in N deplete and replete conditions. An overview of the expression of these 15 *NCRs* for every 4 hour interval can be seen in Figure 6.1.3-1.

All of the *NCRs* with the AGATATTT element in their promoters showed expression variation from 0 to 48 hours in both deplete and replete N conditions. However, NCRs in deplete N seemed not follow any particular expression pattern (Figure 6.1.3-1 A) or exhibit correlation with the expression pattern of Medicago *CCA1/LHY* (*Medtr7g118330*) or *RVE1* (*Medtr5g076960*) genes. In N replete conditions there was a different expression pattern. A peak expression of *RVE1* ortholog *Medtr5g076960* was very similar to the peak point of CCA1 ortholog *Medtr7g118330* just before dawn at 20 hours. While *M. truncatula* circadian ortholog genes *Medtr7g118330* and *Medtr5g076960* did not show any damping towards the end of the time series, the15 NCRs with AGATATTT element in N repletion did dampen (Figure 6.1.3-1 B), as observed earlier (Figure 6.1.2-2 A, B).

The expression profile of 15 NCRs with AGATATTT element was compared with 52 *NCRs* with no CCA1 or RVE1 motif sites. However, these 52 *NCRs* without CCA1 or RVE1 motif sites contained one or the other 5 motifs we presented earlier (Figure 6.1.3-2). Surprisingly, all of these *NCRs* also showed expression variation over time (Figure 6.1.3-2) but with opposite expression response in N depletion (Figure 6.1.3-2, A,C) and repletion (Figure 6.1.3-2, B, D)

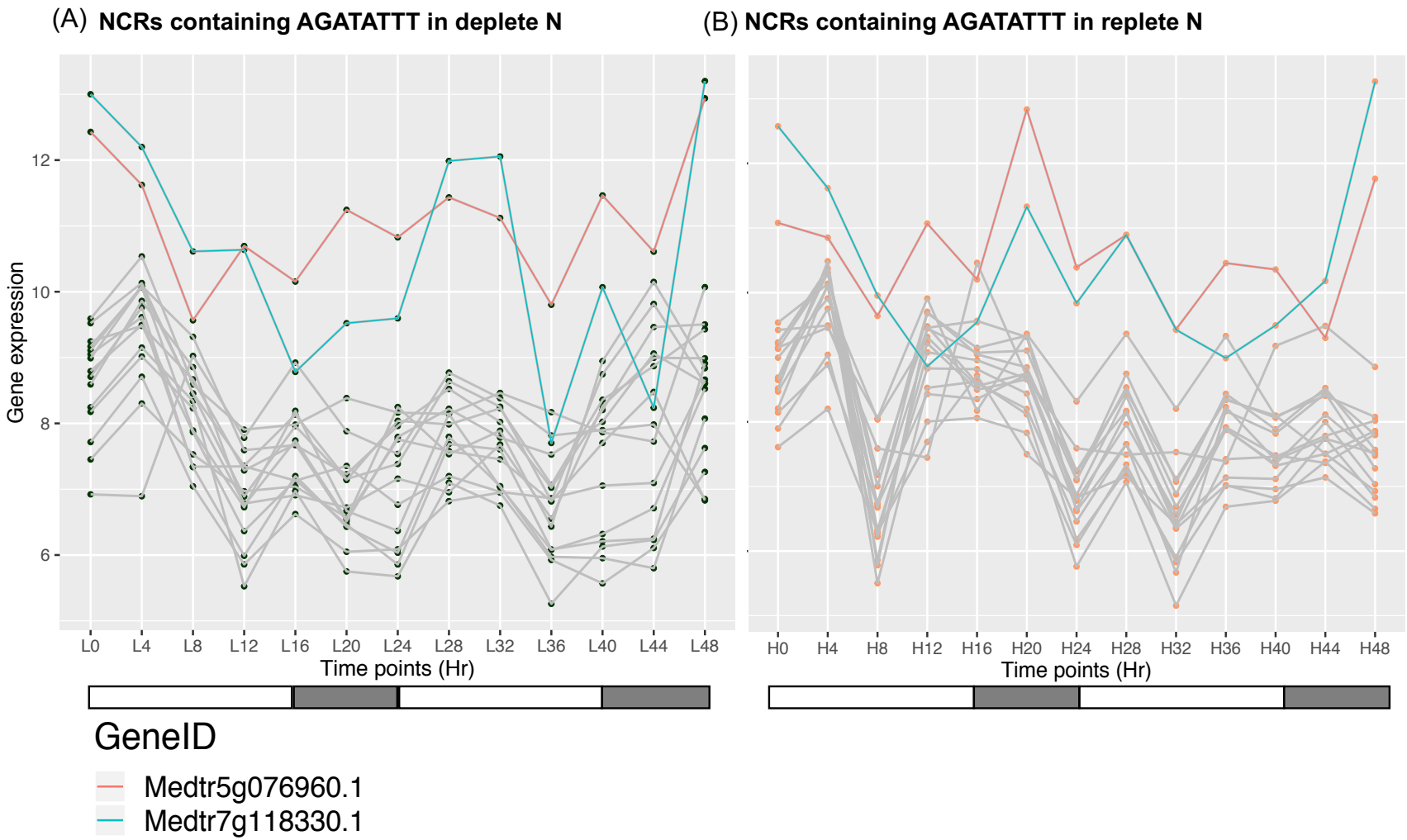


Figure 6.1.3-1: Expression pattern of 15 NCR genes containing CCA1and RVE1 binding sites AGATATTT over time. As well as the NCRs (grey lines), expression of RVE1 ortholog (*Medtr5g076960*) and CCA1 ortholog (*Medtr7g118330*) are denoted by red and blue lines respectively. Unfilled rectangle box below X-axes represent 16 hours of light and grey solid box represent 8 hours of dark. (A) L0-L48 represent time points in hours at deplete N (0.1 mM) conditions. (B) H0-H48 represent time points in hours at replete N (5 mM).

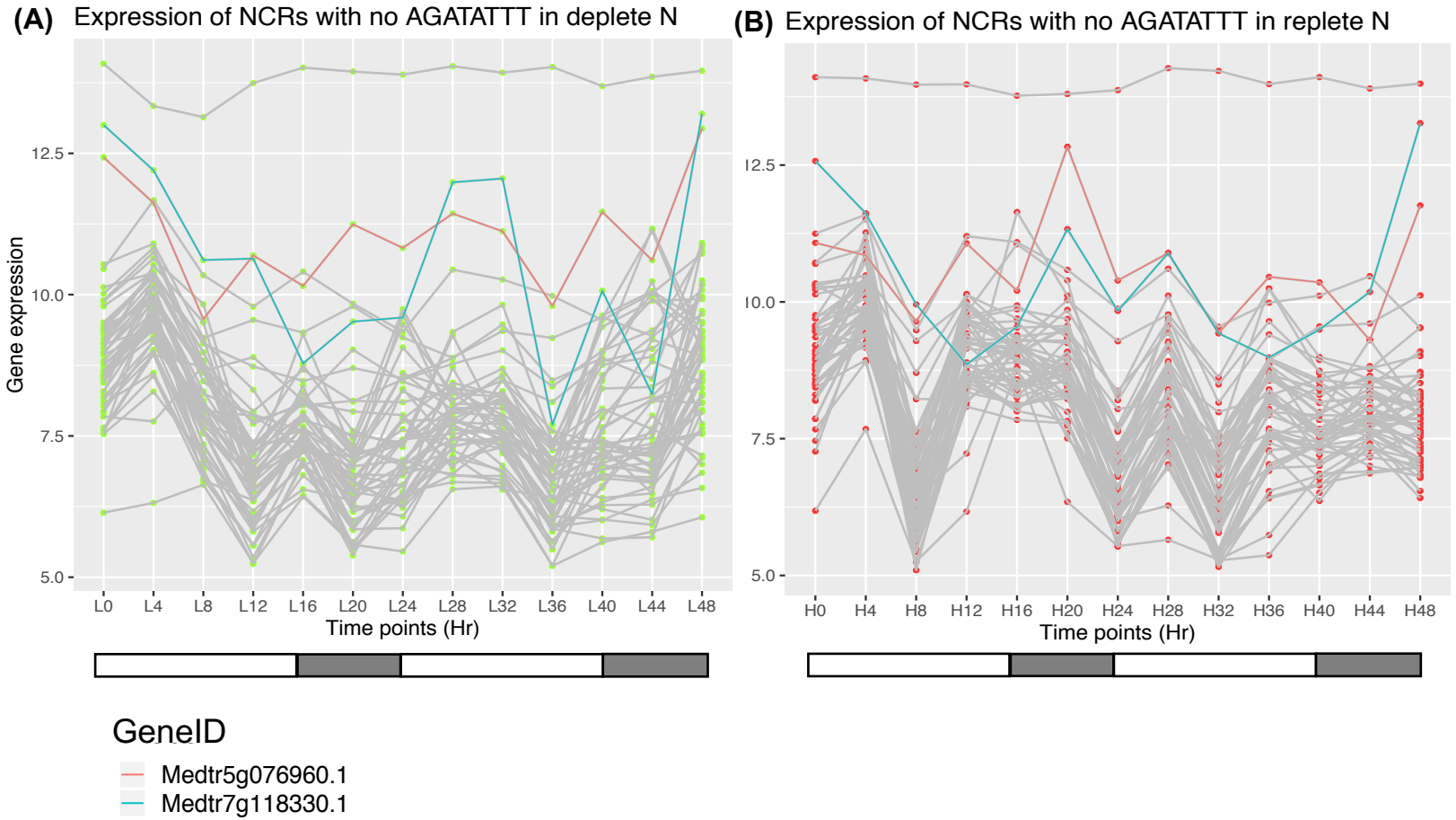


Figure 6.1.3-2: Expression pattern of 15 NCR genes with no CCA1and RVE1 binding sites AGATATTT. As well as the NCRs (grey lines), expression of RVE1 ortholog (*Medtr5g076960*) and CCA1 ortholog (*Medtr7g118330*) are denoted by red and blue lines respectively. Unfilled rectangle box below X-axes represent 16 hours of light and grey solid box represent 8 hours of dark. (A) L0-L48 represent time points in hours at deplete N (0.1 mM) conditions. (B) H0-H48 represent time points in hours at replete N (5 mM).

6.2 Discussion

Nearly all stages of plant development and growth metabolism are under the regulation of light and the clock (de Montaigu et al., 2010; López-Juez and Devlin, 2008; Yakir et al., 2007). The endogenous circadian clock regulates various processes, from photoperiodic control of flowering (Shim et al., 2017; Song et al., 2010), defense response against biotic attack (Hevia et al., 2015; Wang et al., 2011), nutrient acquisition for plant homeostasis (Haydon et al., 2015) to chromatin changes (Nakahata et al., 2007; Ripperger and Merrow, 2011). It is therefore possible that under the natural growth conditions of day and night cycles, the circadian rhythm regulates nodulation and nitrogen fixation in legumes.

Early studies using acetylene reduction assays have been used to show the diurnal changes in N fixation (Minchin and Pate, 1974). However, with only a few studies on soybean and phaseolus (common bean) that have suggested the possible existence of a nodule circadian clock, the importance of circadian rhythms in the control of the rhizobium-legume symbiosis in Medicago is not known. It does however seem likely that, if gene expression in the root is under circadian control, then gene expression in nodules would also be so, since nodule growth, symbiotic N fixation and N assimilation by legumes are all tightly linked to photosynthate C allocation from the plant.

In a study on soybean, *SAT1* (*Symbiotic Ammonium Transporter 1*) was shown to encode a basic helix loop TF with nocturnal expression that was linked to circadian clock genes and important for nodule growth and NH_4 transport (Chiasson et al., 2014). However, the likely existence of nodule circadian clock still remains a complex hypothesis to be proven. While this investigation put special emphasis for the regulation of *NCRs* under circadian control, the role of endogenous phyohormone signalling such as from cytokinin and auxin (Frugier et al., 2008) is also likely to have a strong regulatory effect and should be investigated in future work.

In order to ask if the *NCRs* were under putative circadian control, their expression was analysed over a time series. *NCRs* with CCA1/RVE1-binding

sites in their promoters were found to oscillate over time. However, NCRs lacking these motifs were similarly found to oscillate, not supporting the nodulation clock hypothesis. Yet, the over-representation of CCA1 and RVE1 binding sites on *NCR* promoters is puzzling. It could be that the expression of all *NCRs* is under circadian control, but only some are directly regulated (those with the CCA1/RVE1 motif). Further work is needed here to explore the hypotheses, including complete analyses of the time series transcriptome dataset, analysis of the *NCRs* over time in continuous light, and analysis of *NCR* expression in the *cca1* and *rve1* mutant lines that this thesis has isolated.

Following the use of time series analyses of nodulation at varying levels of N to identify the interaction of NCR genes and circadian clock TFs such as CCA1/RVE1 and ATHB15, these should be investigated with biochemical methods. A yeast-1-hybrid assay, electrophoretic mobility shift assay (EMSA) or high throughput ChIP-sequencing to map the TF protein DNA interaction would be very informative.

NimbleGen microarray probes do not cover all the Medicago genes in the recently annotated genome, thus, to generate complete transcriptomic information, RNA-seq should be performed. Expression of NCR genes in Medicago clock mutants under constant light or dark conditions could also be studied to examine arrhythmic expression or phase shift in the circadian cycle. This work presents a preliminary analysis that has hypothesised a nodule circadian system under the regulatory control of NCR gene expression driven by the presence of CCA1/RVE1 binding sites.

Chapter 7: General Discussion and Future perspectives

Nitrate as a signal molecule regulates plant growth, development, metabolism and gene expression changes (Ruffel et al., 2008, Vidal and Gutiérrez, 2008, Bouguyon et al., 2012). Responses to available nitrogen status occur at the cellular level and are tissue specific. These localised responses in combination lead to systemic regulation of metabolic, physiological and developmental processes that alter plant root system architecture (RSA). However, despite the major advances made to understand local and systemic regulation of N, the full extent of the of NO_3^- signalling pathway is still unknown. In legumes, RSA is also affected by the formation of nodules during plant-rhizobia interactions. While most of the nitrogen uptake and assimilation machinery has been widely studied in Arabidopsis, the regulatory mechanism underlying the uncoupling of NO_3^- assimilation and nodule specific activity in legumes have not been fully investigated and controversial (Cabeza et al., 2014). To address this, investigation of the underlying mechanisms of nitrate effects on nodules in legume RSA is crucial.

7.1 Systems biology approach in *M. truncatula* interaction with N and rhizobia

Nitrate strongly inhibits legume nodule formation (Ferguson et al., 2010) as well as the activity of existing nodules (Vessey and Waterer, 1992). For example, nodules start to senescece within a few days of nitrate exposure (Matamoros et al., 1999). The molecular mechanism of how NO_3^- inhibits nodule formation has been reported to resemble that of autoregulation of nodulation (Reid et al., 2011). However, the exact mechanism of nitrate-induced inhibition remains poorly understood, although this could be a policing mechanism to limit the nodule formation under conditions with sufficient nitrate availability. Such nitrate-dependent regulation of nodule formation is illustrated by the existence of mutants that hypernodulate even in the presence of high nitrate levels. This indicates that nitrate is not the inhibiting factor itself, but that

it could lead to long distance signaling and suppression of nodule activity (Searle et al., 2003).

Using a systems biology approach we utilised the *Medicago truncatula* autoregulation of nodulation mutant *sunn-1* (disrupted in long-distance signalling AON) and A17 wild-type to investigate root architecture studies during nodulation. We integrated our root phenotypic studies with microarray expression profiling to investigate the effect of rhizobium and N in *M. truncatula*. In this study we focused on how LR development and nodule formation were balanced in *M. truncatula* according to the external N level. This enabled us to study the effect of rhizobia and nitrate levels (deplete or replete) in the development of Medicago root system architecture.

Our phenotypic data showed that both A17 and *sunn-1* were found to have a shorter primary root length when inoculated with rhizobium, independent of N availability. Both the genotypes A17 and *sunn-1* were demonstrated to show different N-status-dependent phenotypic responses when seedlings were inoculated with rhizobium. This response was more enhanced in *sunn-1* at replete N with a significantly ($P<0.05$) higher number of nodules and LRs than A17 whether rhizobia is present or absent. The hypernodulant *sunn-1* with more number of LR development but shorter LR length at deplete N indicates that AON pathway components could be involved in balancing LR development. This is in line with the from the previous studies that AON pathway genes could be candidates for regulating LR development in response to N (Jin et al., 2012). Hence, our result indicate a complex regulatory balance between nodule development and RSA , with N and rhizobium inoculation to have a partially additive effect.

To study this N-dependent balance between LR and nodule development, microarray expression profiling was performed in A17 and *sunn-1*. The microarray experiment enabled us to identify the underlying genes observed in our phenotypic studies that control root architecture response to N and rhizobia. Out of the 6,910 genes identified as differentially expressed on our microarray expression data, rhizobia-regulated responses were stronger in both A17 and *sunn-1* ($FC > 1.5$ and $FC < -1.5$) irrespective of N status (Figure

3.1.3 A, B). However, *sunn-1* showed more number of rhizobia-responsive genes than in A17 (FC > 1.5 and FC < -1.5) at N repletion (Figure 3.1.3 B).

From the hierarchical clustering of the transcriptome, three major distinctive clusters (4, 6 and 9) have been shown with the largest number of DEGs. Cluster 4 represented genes were involved mainly in the nodulation pathway, showing a rhizobia effect. Cluster 6 comprised N-related genes that had an N effect, while cluster 9 represented a genotype effect related to *sunn-1*. Out of these three clusters, we have presented cluster 4 which includes *NCR* gene family of Medicago in more detail. In contrast to the findings of *NCR* gene expression only in symbiotic nodules (Guefrachi et al., 2014), our microarray data revealed *NCR* gene expression in roots and roots in early stages of nodulation, with NCRs being expressed alongside nodulation regulatory genes,

This enabled to identify candidate genes that control the responses and cross talk between N and rhizobium in *M. truncatula* roots symbiosis. Such analysis of the regulatory genes and processes provides a better understanding of how legumes can balance N uptake, assimilation, LRs development during the nodulation symbiosis. This suggests a tight coregulation of LR development and nodule formation,

7.2 Transcription factor identification from the promoter analysis and motif conservation

To explore the gene regulatory pathways that could underlie the possible co-regulatory balance of LR development and nodule formation, we analysed the promoter region. Use of MEME motif searching in the differential expression groups of N and rhizobia responsive gene promoters led to the identification of conserved motif sites (TCATGAAAGGTT, TATAAAGTGATCA, CAACACATTGAT, AGAGACATTTAA, TTTTACAACTCC). From the multiple sequence analysis we revealed the landscape, homology and conservation of these motifs in a representative NCR promoter. This work implicated significant motifs as PSSMs that could serve as putative cis-regulatory binding sites for TF(s) to regulate NCR gene expression of *Medicago*.

This work presents a preliminary hypothesis that the nodule circadian system controls NCR genes via the presence of CCA1/RVE1 motif sites in their

promoters. The CCA1/RVE1 binding sites are overrepresented in the NCR promoters and the presence of the core GATA element is likely to enable circadian regulation. GATA transcription factors bind to a consensus DNA sequence (A/T)GATA(A/G) and are involved in circadian- regulated gene expression and N metabolism (Manfield et al., 2007; Reyes et al., 2004). In *Arabidopsis* several GATA transcription factor proteins are found to interact with cis-acting elements of light responsive promoter genes (Jeong and Shih, 2003; Teakle et al., 2002). Finding of these GATA cis-elements in promoters of NCR genes that have highly nodule-specific expression (Guefrachi et al., 2014), are in line with the findings from (Zhang et al., 2015) in which real-time PCR profiling of soybean TF gene *GmGATA* revealed its tissue specific expression in stem, young leaf, flower including roots.

Another possible influence of the core GATA cis-element on the CCA1 motif AGAGA(C/T)ATTTAA could be related to the nitrogen fixation function *Medicago* root nodule and N metabolism. GATA transcription factors have been implicated in controlling the N metabolism pathway in *Arabidopsis* and rice (Reyes et al., 2004) and in fungi *Neurospora crassa* (Fu and Marzluf, 1990). Several studies in Arabidopsis have identified GATA motifs in the regulatory region of nitrate assimilation genes (Oliveira and Coruzzi, 1999; Rastogi et al., 1997).

Given the NCR gene sequence diversity, our findings of just 27 NCR gene promoters with the core GATA element are in contrast to the wide distribution of GATA motifs in the target promoter sequence from a study in soybean (Zhang et al., 2015). Aside from circadian regulation and N metabolism, it should be noted that GATA motifs also enable regulation of plant development (Shikata et al., 2004; Zhao et al., 2004). Thus, our identification of NCR promoters with conserved CCA1 motifs containing the core GATA element supports the multiple reports of GATA motifs in other plants. GATA motifs have also been reported in other legumes including soybean (Zhang et al., 2015) and *Medicago sativa* (a close relative of *M. truncatula*) (Fugate et al., 2014). As such our findings of CCA1 and GATA motifs in NCR family promoters in Medicago represent a novel finding to add to the current knowledge of GATA factor family.

In this work, putative TF ortholog genes in *Medicago* were identified that could act as potential binders to the *de novo* discovered conserved sites in *NCR* promoters. Based on orthology analysis, the plant developmental protein ATHB15/ATHB16 that we found as putative TF binders are other candidates to support the role of NCR gene family beyond plant defense antimicrobial. ATHB15 has been demonstrated to be involved in vascular development through miR166-mediated ATHB15 mRNA regulation (Kim et al., 2005). In Arabidopsis, ATHB15 has been shown to be involved in procambial function of leaves and roots, and is an important transcriptional regulator responsible for early vascular development (Ohashi-Ito and Fukuda, 2003).

Mutagenised populations serve as an indispensable resources and mutant utillisation has been a useful approach for gene functional studies through forward and reverse genetics (Bolon et al., 2011; Wu et al., 2005). Aside from transcriptomic investigation, by screening and isolation of the mutants from the Tnt population of NF17115, NF16461, NF13921, NF5993, NF5999 and NF0908, we have shown the utility of the Medicago Tnt mutant. The resources utilised in the mutant analyses will provide the basis for future screening of genes of interest from the Tnt Medicago mutant population. For example, progenies from the homozygous mutant line NF17115, NF13921 and NF5993 can be used for further gene function studies. These mutant characterisation studies will support to establish additional roles of NCR gene family and its possible regulation under circadian control. If confirmation of the loss of gene function in *Medtr3g109800* is confirmed, the NF5993 homozygous line will be a promising candidate to analyse in more detail to investigate the interaction between nodule development and root development.

Another line of study is the comparative phenotypic characterisation of *cca1, rve1* and *athb15* mutant plants. As highlighted in Chapter 5, our preliminary phenotypic observation (data not shown) of these homozygous mutant lines showed different root development and growth timing with *cca1* and *rve1* putative mutant lines NF17115, NF16461 and NF13921 showing longer time requirement of > 2 weeks for complete root development, compared to WT-R108, WT-A127 and *athb15* mutant lines that only took 7 days. *cca1*

and *rve1* homozygous mutant lines analysed under the same growth conditions (16 hrs light 8 hrs dark, at 25ºC) also showed a reduction in growth rate.

Molecular characterisation was performed in the Medicago mutant population developed by Tnt retrotransposon insertions. Multiple insertions copies of Tnt1 can present each line (Tadege et al., 2008) despite its stability from seed to seed (d'Erfurth et al., 2003). This necessitates removal of the extraneous alleles that arose from random insertions in gene coding region (Tadege et al., 2008) via back-crossing. Nevertheless, as future line of work to make use of Medicago mutants generated by the fast neutron bombardment method, our Medicago gene query sequence used for BLAST search can be revised. This is because the FNB probe sequence in the FNB mutant database was found to be very short. This could suggest that the same query sequence used for Tnt mutant search was not applicable for FNB mutant search.

The IRLC legume clade including *Medicago truncatula*, *Pisum sativum* and *Vicia faba* form indeterminate nodules when housing symbiotic rhizobium bacteria. Bacteroid differentiation inside the symbiotic nodule cells in such IRLC clade has so far been reported to be controlled by a large class of cysteine rich peptides, NCRs. A notable feature of the *NCR* family in *Medicago truncatula* is that it is very large, comprising about 600 genes whose expression is specific to nodules only (Mergaert et al., 2003; Nallu et al., 2013).

Classification of NCRs has been reported to be difficult due to the highly diverse amino acid sequences of genes in the family (Nallu et al., 2014). A previous hypothesis suggested that they could have evolved from defensin like anti-microbial peptides, regulating the endo-symbiotic bacteria to undergo terminal differentiation and thereby controlling symbiotic efficiency (Farkas et al., 2014; Kereszt et al., 2011; Van de Velde et al., 2010). However, our and other transcriptome analyses have shown that NCRs have distinct spatio-temporal profile with *NCR* genes expressed in successive waves at different nodule zones, not simply consistent with a defensin-like role (Guefrachi et al., 2014; Roux et al., 2014).

Studies on cysteine rich peptides indicated that they can also function as signalling peptides regulating the root developmental pathway (Marshall et al., 2011). Recent findings (Wang et al., 2017; Yang et al., 2017) have also

speculated the existence of a dual nature for NCRs, possessing both pro-symbiotic and anti-symbiotic properties. This is based on the evidence that *NITROGEN FIXATION SPECIFICITY* (*NSF*) genes *NSF1* and *NSF2* encode for NCRs that functions as negative regulator of symbiotic efficiency. Therefore, there is accumulating evidence that the NCRs have additional roles beyond just an anti-microbial function. This is in line with the presence of putative ATHB and circadian clock CCA1/RVE1 transcription factor binding sites on NCR promoters that can possibly regulate NCR gene expression. The presence of the highly conserved 4-6 cysteine residues in the NCRs is another interesting feature that has not been fully explored.

Diverse spatio-temporal expression profiles (Nallu et al., 2014), high specificity across nodules (Mergaert et al., 2003; Nallu et al., 2013), amino acid sequence variation leading to varying isoelectric points from 3.2 to 11.25 (Kondorosi et al., 2013), and the presence of cysteine rich residues that can act in signalling pathways (Marshall et al., 2011) all point towards the ability of NCRs to play diverse roles as the family expanded. This work has shown that the Nodule Cysteine rich peptides could have a role in controlling the regulation of nodule development under different N levels and with regulatory input from the circadian system.

7.3 Circadian rhythm of NCR gene family and symbiosis

Aside from the TF Medicago orthologs that we found, we have started to ask if *NCRs* oscillate over time under the possible control of CCA1 and RVE1 transcription factors. Our time-series transcriptomic data in deplete (0.1 mM, N) and replete (5 mM, N) aimed to investigate NCR regulation under the combinatorial action of circadian rhythm and N level.

The existence of a nodule expressed circadian clock was proposed from transcriptomic profiling of soybean *GmbHLHm1* (Chiasson et al., 2014) and common bean (Dalla Via et al., 2015). Transcriptomic changes were also studied in the interaction between soybean-*Bradyrhizobium japonicum* (Chiasson et al., 2014) and Phaseolus-*Rhizobium etli* (Dalla Via et al., 2015). Hence, it would be very interesting to ask if there is any circadian-N regulation in these plant-mutualist systems.

It is known that circadian rhythms play a crucial role in anticipation of environmental changes, thereby affecting development and stress responses. Many genes have been found to be clock-regulated (Harmer, 2009; Michael and McClung, 2003; Pruneda-Paz and Kay, 2010). Yet, the integration of over-ground signals to the root by clock has not been well defined. How the circadian system in roots are adjusted after rhizobium entry and for nodule organogenesis poses a challenging investigation for which this work offers some novel insights.

Nitrogen fixation in legumes has been demonstrated to be light and temperature responsive under the cycles of day/night. For example, red light was shown to have no effect on rhizobial growth, while blue light was found to inhibit nodulation in *Lotus japonicus* roots that were inoculated with *Mesorhizobium loti* (Shimomura et al., 2016). However, the interaction between circadian rhythms and symbiosis have not been fully explored.

In Chapter 6, we found that NCRs in deplete N conditions demonstrated a wide variation of expression levels as compared to replete N but all appeared to have some oscillatory variation (Figure 6.1.2-2, C and D). In N replete conditions, NCRs expression showed damping towards the end of the 48 hour period, relative to their expression at 0 hour (Figure 6.1.2-2, A and B). All of the *NCRs* with or without AGATATTT element in their promoters showed expression variation from 0 to 48 hours in both deplete and replete N conditions. However, expression response in N depletion (Figure 6.1.3-2, A,C) and repletion (Figure 6.1.3-2, B, D) was opposite.

This observation could be related with the complex defense responses in host plants that are tightly linked with the life cycles of biotrophic pathogens. Such two way signaling between host plants and microbes is finely modulated by the day/night cycle. Hence, the host-microbe interaction may be dictated by the circadian clock of host. In plants, this link has been studied in *Arabidopsis* defense responses against *Hyaloperonospora arabidopsidis* (Wang et al., 2011) and *Pseudomonas syringae (*Bhardwaj et al., 2011). Host-plant roots must differentiate symbiont partner from pathogen (Oldroyd, 2013), to form successful nodules and the control of bacteroid population is likely to be linked

to the day/night cycle. Further analyses of our transcriptome data could reveal a circadian-symbiosis interplay in a host plant-controlled mechanism.

A model of circadian clock control of plant innate immunity has been based on circadian-regulation of defense gene expression. Although all of the NCR targets are till date unknown, some NCRs (such as NCR247 and NCR211) functional association targeting bacteria in symbiotic cells have been shown (Kim et al., 2015; Penterman et al., 2014; Tiricz et al., 2013). This is in line with previous reports of NCRs having a resemblance to defensin type antimicrobial peptides (Nallu et al., 2013) that are a part of the plant innate immune response (Maroti et al., 2011). Collectively, these findings on just a few NCRs, suggesting a role in host immune responses, in combination the putative presence of CCA1 binding sites in NCR promoters lead to the hypothesis that there is circadian regulation to time the immune suppression, enabling successful symbiosis.

While circadian rhythms have so far been well-described in host plants including *Arabidopsis*, the circadian effect on rhizosphere community has been less well studied. The bacterial community in the soil, including *Burkholderiaceae*, *Rhodospirillaceae*, *Planctomycetaceae* and *Gaiellaceae* have been reported to fluctuate and respond to circadian rhythms that in turn exert influences on/by the host plant (Staley et al., 2017). Hence, such findings indicate another interesting future perspective to investigate the legume-rhizobium symbiosis under the circadian control with the inoculation of different rhizobial strains.

Whether a nodulation circadian clock exists or not still poses a challenging hypothesis to prove. In our study we have hypothesized that it could be similar to the clock in roots and that N fixation in legume-symbiosis could be circadian-regulated due to the enrichment of CCA1/RVE1 binding sites on nodule-specific gene family, NCRs. Our result in Chapter 6 presented the preliminary transcriptomic time series data of NCR gene family with AGATATT and core GATA motif. However, the regulatory effect of phytohormones, especially auxin and cytokinin for circadian regulation, nodule organogenesis and root development in addition cannot be understated. The circadian clock has been linked to the global regulation of endogenous auxin signaling

(Covington et al., 2008)., altered cytokinin response affects the circadian oscillations in *Arabidopsis* (Zheng et al., 2006) and there is a direct link of cytokinin to nodule organogenesis in *Medicago* (Ariel et al., 2012). This highlights the importance of investigation to study the effect of phytohormonal signalling in nodule circadian system.

In symbiotic relationships, the association between plant and symbiont microbe is highly balanced and host plants seem to control the degree of colonisation. Mutualistic symbionts, just like pathogens, have been shown in various studies to activate immune response in host plants upon perception (Bonfante and Genre, 2010; Cao et al., 2017). For successful symbiosis benefitting both the partners, the symbiont microbe escapes or evades the host surveillance mechanism of immune response to successfully establish efficient symbiosis (Liu et al., 2018; Navarro et al., 2008; Schafer et al., 2009).

In addition, the role of phytohormones in promoting nodule development and symbiosis is well known. For example, cytokinin (Heckmann et al., 2011; Reid et al., 2016), strigolactones (Breakspear et al., 2014; De Cuyper et al., 2015) and local accumulation of auxin (Roy et al., 2017; Suzaki et al., 2012) not only exert a positive role by promoting nodule organogenesis but also regulate gene expression in Medicago nodules. Apart from the positive regulation, other phytohormones such as ethylene (Oldroyd et al., 2001), abscisic (Ding et al., 2008) and gibberellic acid (Ferguson et al., 2005) exert negative regulation of infection thread formation and nodule development. Hence, the critical role of phytohormones in nodule organogenesis and thier spatio-temporal regulation in root nodules are important to investigate in the light of cross-talk with circadian rhythm.

Despite the studies on phytohormonal regulation of root nodule symbiosis, we still lack insight into the molecular mechanisms that underlie the complex cross-talk between nodule development and spatio-temporal expression of nodule specific genes. If phytohormonal effects can be considered with further investigation of nodule circadian rhythms, such studies will provide a better understanding of how circadian regulation interacts with auxins, cytokinin and ABA to regulate the timing of nodulation and expression of nodule specific genes.

Another important aspect of rhythmic metabolism in plant circadian system is the clock coordination of carbon metabolism. Interestingly, carbon metabolism and mobilisation of photosynthates that drives plant development are clock-controlled. This has been well-studied in Arabidopsis roots where the circadian clock is set by sucrose, a photosynthesis-related signal that leads to altered root rhythmicity (James et al., 2008). (Haydon et al., 2013) reported that sucrose levels peak in the morning as a 'metabolic dawn' due to decreased PRR7 (pseudo response regulator 7) expression, which in turn represses CCA1 transcription to set the clock. Thus the circadian clock has a dual interaction in regulating and being regulated by carbon metabolism to adjust clock output and modulate plant growth. While such interactions have been widely studied in Arabidopsis, clock genes appear to be conserved across species, including legumes. The possible circadian regulatory effect on nodule organogenesis, rhizobium symbiosis via the overrepresentation of CCA1 binding sites in promoters of the nodule specific NCR gene family is an interesting specific example of this conservation.

Related to this, N fixation in root nodules in exchange for photosynthate carbon is a costly process for the legume plants. As such, carbon requirements become higher for atmospheric N_2 fixation and to prevent C loss, host legumes can discriminate against ineffective rhizobial strains with inefficient nitrogen fixation, by two mechanisms; partner choice or sanctions (Heath and Tiffin, 2009). This indicates that the host legume can distinguish N fixation levels and accordingly allocate photosynthate C to individual nodules. This regulation of resource allocation was best exemplified from a study whereby severity of sanction response was increased depending on external nitrate availability (Kiers et al., 2003). However, the mechanism of when legume recognises inefficient rhizobia in the nodule and when sanction occurs remain largely elusive.

It has been demonstrated that differentiated bacteroids in root nodule confer net benefits to the host legume due to their more efficient nitrogen fixation, as compared to non-differentiated bacteroids (Oono and Denison, 2010; Oono et al., 2010). *Medicago truncatula* with indeterminate root nodule type accommodates such differentiated bacteroids, indicating a higher return

of fixed N at the expense of C allocation for nodule formation. In conjunction with this, NCRs controlling discrimination against incompatible microsymbionts (Yang et al., 2017) and NCR mediation to optimise bacteroid over proliferation and N fixation process (Van de Velde et al., 2010) have both been found. Taken together with the presence of CCA1 binding sites in NCR promoter regions, it is therefore possible that NCRs may have additional roles in facilitating efficient N fixation which involve C allocation, under the regulation of circadian system.

As highlighted earlier, evidence for clock regulation of phytohormones and metabolites is accumulating. A prominent feature of clock regulation of metabolites is the mechanism that the circadian clock regulates C and N metabolism depending on their metabolite status, and in turn the metabolites themselves can act as inputs for the clock (McClung and Gutierrez, 2010). More recently, NCRs have also been demonstrated to exhibit functions enabling symbiont membrane permeability for metabolite exchange (Mergaert et al., 2017). This indicates that the host legume metabolic network and rhizobium are perhaps interlinked for metabolite exchange to fulfil the diverse functions of cysteine rich peptides such as NCRs. In line with our study, significant enrichment of the CCA1 binding sites on NCR promoter region provides an intriguing premise to investigate how the circadian clock can co-ordinate the large and nodule-specific NCR gene family. Emerging data shows that circadian clock also functions to enhances survival and biomass accumulation (Dodd et al., 2005; Green et al., 2002). With such evidence, the study of CCA1/LHY interactions with Medicago NCR genes can yield insights into regulatory effects on fitness, symbiotic performance and nodule growth.

Determining how the master clock components such as CCA1 control their targets and output in a tissue specific manner by integrating metabolite, phytohormonal pathway, external nutrient availability and environmental cues from microbes enable understanding of how biological rhythms underpin growth, fitness and symbiosis. As such, systems approaches with genome-wide investigation to study transcriptome dynamics and gene expression levels such as microarray, NGS technologies and metabolomics will continue to serve as crucial means to characterise circadian system. However, detailed

information on physical and regulatory interactions is necessary for a better understanding to characterise the existence of a nodule specific circadian clock and symbiosis.

Conclusion

Consistent with the assertion that major cell types in *A. thaliana* possesses independent circadian oscillator (Thain *et al.*, 2000), It is possible that there is specialisation of circadian clock function in specific cells (Xu *et al.*, 2007) including root cells (James *et al.*, 2008) and vascular tissue (Para *et al.*, 2007). With lateral roots that have been demonstrated to be under the circadian rhythm (Voss et al., 2015), accumulating evidence in circadian control of plant-microbe interaction and rhizosphere community as described before, the fact that how circadian system will not subtly affect nodule organogenesis and rhizobium-legume symbiosis is questionable.

In addition to time series analyses of transcriptome to validate our *in silico* findings of NCR gene interaction with circadian clock TFs, *in vitro* analysis of yeast-1-hybrid assays or EMSA can be performed. As a future line of work, next generation sequencing techniques such as RNA-sequencing and high throughput ChIP-sequencing can be utilised to generate a reliable transcriptomic data and to establish genome wide protein-DNA interaction.

Expression of NCR genes in Medicago clock mutants with inoculation of different rhizobial mutant strains can be experimented under natural, constant light or dark conditions to examine arrhythmic expression or clock mutant phenotype that can affect nodulation. As such NCR gene expression regulations in these conditions can be analysed. Overall, our findings of conserved motifs in NCR promoters and its putative TFs orthologs support the multiple roles NCR gene family are thought to play.

Taken together, our data support the model of rapidly evolving NCR gene family for the gain of new function in root symbioses and nodulation. As such, functional genomics study in the mutants together with transcriptomic time series analysis of NCRs and its regulatory control under clock motif sites will help us in elucidating additional roles of NCRs. This supports the possible and multiple functions of NCRs as DEFLs have been reported to play dual roles in defense and developmental plant signalling during plant-microbial interaction.

Appendices

Table 1: Gene annotation term enrichment and their p-values for DEGs in 11 clusters from the A17 and *sunn-1* microarray analysis.

Cluster	Number of DEG in each cluster	Number of genes annotated with calcium	P-value for calcium term enrichment	Number of genes annotated with auxin	P-value for auxin term enrichment
1	1	0	1	0	1
2	2	0	1	0	1
3	3	0	1	0	1
4	1283	5	0.92215585	6	0.989739
5	249	2	0.46245468	1	0.9123041
6	3137	28	0.00710738	37	0.0524881
7	118	2	0.16667457	1	0.6808915
8	190	3	0.11366497	1	0.8426073
9	1896	3	0.99985516	20	0.3436502
10	24	0	1	0	1
11	7	0	1	0	1
Total	6910	43		66	

Cluster	Number of DEG in each cluster	Number of genes annotated with cytokinin	P-value for cytokinin term enrichment	Number of genes annotated with UDP	P-value for UDP- term enrichment
1	1	0	1	0	1
2	2	0	1	0	1
3	3	0	1	0	1
4	1283	4	0.13071325	7	0.96626694
5	249	2	0.05740628	5	0.07987924
6	3137	4	0.81627659	37	0.03043468
7	118	0	1	0	1
8	190	0	1	3	0.25737446
9	1896	1	0.97072746	12	0.96046804
10	24	0	1	0	1
11	7	0	1	0	1
Total	6910	11		64	

Table 1 continued

Cluster	Number of DEG in each cluster	Number of genes annotated with nitrogen	P-value for nitrogen term enrichment	Number of genes annotated with transport	P-value for transport term enrichment
1	1	0	1	0	1
2	2	0	1	0	1
3	3	0	1	0	1
4	1283	11	0.57134017	57	0.082390988
5	249	1	0.8904783	10	0.453017033
6	3137	35	0.02954066	142	0.00097173
7	118	0	1	1	0.989212257
8	190	2	0.49474321	5	0.844264719
9	1896	11	0.96312179	43	0.999987666
10	24	0	1	0	1
11	7	0	1	0	1
Total	6910	60		258	

Cluster	Number of DEG in each cluster	P-value for transport term enrichment	Number of genes annotated with kinase	P-value for kinase term enrichment	Number of genes annotated with redox
1	1	1	0	1	0
2	2	1	0	1	0
3	3	1	0	1	0
4	1283	0.082390988	49	0.9999991	3
5	249	0.453017033	16	0.5790397	0
6	3137	0.00097173	276	9.96221E-	26
7	118	0.989212257	1	12	2
8	190	0.844264719	18	0.9996991	0
9	1896	0.999987666	95	0.07471773	5
10	24	1	0	0.999638	0
11	7	1	0	1	0
Total	6910		455		36

Table 1 continued

Cluster	Number of DEG in each cluster	Number of genes annotated with cysteine	P-value for cysteine term enrichment	Number of genes annotated with flavonoid	P-value for flavonoid term enrichment
1	1	0	1	0	1
2	2	0	1	0	1
3	3	0	1	0	1
4	1283	190	7.86342E-85	1	0.91515242
5	249	1		2	0.06728991
6	3137	47	0.9999448	8	0.11701462
7	118	3	1	0	1
8	190	5	0.8313023	1	0.28455329
9	1896	16	0.8536207	0	1
10	24	0	1	0	1
11	7	0	1	0	1
Total	6910	262		12	

Table 2: NCR probe specificity check on NimbleGen microarray by BLASTN for randomly selected probe ID to represent the BLAST results. 530 probes were found on NimbleGen with 2-3 probes represented for a single gene. Few of the probe IDs with their alignments to *M.truncatula* genome have been given. Probes of ~45-60 bp were found to have 100% match identity to that of the target NCR genes. However, despite significant alignments and 100% homology few of the probe sequence were not annotated as NCRs. E.g probe sequence of Medtr1g042910. The top hit was confirmed to be a match for NCRs. The second top hit typically only aligned with <20bp identity and with higher E-value than the first top hit.

Sequence ID	Probe ID	Overlapping gene	Annotation	Alignment	E-value	% Identity	Alignment length
Medtr1g042200.1	Medtr1g042200.1P00266	*MTR_1g042200*	NCR	60	2.00E-26	100	60
Medtr1g042200.1	Medtr1g042200.1P00675	*MTR_1g042200*	NCR	60	2.00E-26	100	60
Medtr1g042200.1	Medtr1g042200.1P00182	*MTR_1g042200*	NCR	60	2.00E-26	100	60
Medtr3g028380.1	Medtr3g028380.1P00164	*MTR_3g028380*	NCR	56	4.8E-24	100	56
Medtr3g028380.1	Medtr3g028380.1P00099	*MTR_3g028380*	NCR	60	2.00E-26	100	60
Medtr1g042910.1	Medtr1g042910.1P00934	-	-	60	2.00E-26	100	60
Medtr1g042910.1	Medtr1g042910.1P00065	-	-	59	7.8E-26	100	59
Medtr1g042910.1	Medtr1g042910.1P00684	-	-	60	2.00E-26	100	60
Medtr6g055160.1	Medtr6g055160.1P00566	*MTR_6g055160*	NCR	48	2.90E-19	100	48
Medtr6g055160.1	Medtr6g055160.1P00211	-	-	60	2.00E-26	100	60
Medtr6g055160.1	Medtr6g055160.1P00002	-	-	60	2.00E-26	100	60
Medtr8g036850.1	Medtr8g036850.1P00267	-	-	60	2.00E-26	100	60
Medtr8g036850.1	Medtr8g036850.1P00121	*MTR_8g036850*	NCR	60	2.00E-26	100	60

Bibliography

Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Mas, P., and Kay, S.A. (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science 293*, 880-883.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol 215*, 403-410.

Alunni, B., Kevei, Z., Redondo-Nieto, M., Kondorosi, A., Mergaert, P., and Kondorosi, E. (2007). Genomic organization and evolutionary insights on *GRP* and *NCR* genes, two large nodule-specific gene families in *Medicago truncatula*. *Mol Plant Microbe Interact 20*, 1138-1148.

Alvarez, J.M., Vidal, E.A., and Gutierrez, R.A. (2012). Integration of local and systemic signaling pathways for plant N responses. *Curr Opin Plant Biol 15*, 185-191.

Ameline-Torregrosa, C., Wang, B.B., O'Bleness, M.S., Deshpande, S., Zhu, H., Roe, B., Young, N.D., and Cannon, S.B. (2008). Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol 146*, 5-21.

Ariel, F., Brault-Hernandez, M., Laffont, C., Huault, E., Brault, M., Plet, J., Moison, M., Blanchet, S., Ichante, J.L., Chabaud, M.*, et al.* (2012). Two direct targets of cytokinin signaling regulate symbiotic nodulation in *Medicago truncatula*. *Plant Cell 24*, 3838-3852.

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. *Nucleic Acids Res 43*, W39-49.

Benkova, E., and Bielach, A. (2010). Lateral root organogenesis - from cell to organ. *Curr Opin Plant Biol 13*, 677-683.

Benlloch, R., d'Erfurth, I., Ferrandiz, C., Cosson, V., Beltran, J.P., Canas, L.A., Kondorosi, A., Madueno, F., and Ratet, P. (2006). Isolation of mtpim proves Tnt1 a useful reverse genetics tool in *Medicago truncatula* and uncovers new aspects of AP1-like functions in legumes. *Plant Physiol 142*, 972-983.

Bensmihen, S., de Billy, F., and Gough, C. (2011). Contribution of NFP LysM domains to the recognition of Nod factors during the *Medicago truncatula/Sinorhizobium meliloti* symbiosis. *PLoS One 6*, e26114.

Bhardwaj, V., Meier, S., Petersen, L.N., Ingle, R.A., and Roden, L.C. (2011). Defence responses of *Arabidopsis thaliana* to infection by *Pseudomonas syringae* are regulated by the circadian clock. *PLoS One 6*, e26968.

Bolon, Y.T., Haun, W.J., Xu, W.W., Grant, D., Stacey, M.G., Nelson, R.T., Gerhardt, D.J., Jeddeloh, J.A., Stacey, G., Muehlbauer, G.J.*, et al.* (2011). Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. *Plant Physiol 156*, 240-253.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics 19*, 185-193.

Bonfante, P., and Genre, A. (2010). Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nature Communications 1*, 48.

Bordage, S., Sullivan, S., Laird, J., Millar, A.J., and Nimmo, H.G. (2016). Organ specificity in the plant circadian system is explained by different light inputs to the shoot and root clocks. *New Phytol 212*, 136-149.

Bouguyon, E., Gojon, A., and Nacry, P. (2012). Nitrate sensing and signaling in plants. *Semin Cell Dev Biol 23*, 648-654.

Branca, A., Paape, T.D., Zhou, P., Briskine, R., Farmer, A.D., Mudge, J., Bharti, A.K., Woodward, J.E., May, G.D., Gentzbittel, L.*, et al.* (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proc Natl Acad Sci* U S A *108*, E864-870.

Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., Morieri, G., Mysore, K.S., Wen, J., Oldroyd, G.E.*, et al.* (2014). The root hair "infectome" of *Medicago truncatula* uncovers changes in cell cycle genes and reveals a requirement for Auxin signaling in rhizobial infection. *Plant Cell 26*, 4680-4701.

Brewin, N.J. (2004). Plant Cell Wall Remodelling in the *Rhizobium*–Legume Symbiosis. *Critical Reviews* in Plant Sciences *23*, 293-316.

Brown, P., Baxter, L., Hickman, R., Beynon, J., Moore, J.D., and Ott, S. (2013). MEME-LaB: motif analysis in clusters. *Bioinformatics 29*, 1696-1697.

Cabeza, R., Koester, B., Liese, R., Lingner, A., Baumgarten, V., Dirks, J., Salinas-Riester, G., Pommerenke, C., Dittert, K., and Schulze, J. (2014). An RNA sequencing transcriptome analysis reveals novel insights into molecular aspects of the nitrate impact on the nodule activity of *Medicago truncatula*. *Plant Physiol 164*, 400-411.

Canales, J., Moyano, T.C., Villarroel, E., and Gutierrez, R.A. (2014). Systems analysis of transcriptome data provides new hypotheses about *Arabidopsis* root response to nitrate treatments. *Front Plant Sci 5*, 22.

Cannon, S.B., Ilut, D., Farmer, A.D., Maki, S.L., May, G.D., Singer, S.R., and Doyle, J.J. (2010). Polyploidy did not predate the evolution of nodulation in all legumes. *PLoS One 5*, e11630.

Cao, Y., Halane, M.K., Gassmann, W., and Stacey, G. (2017). The Role of Plant Innate Immunity in the Legume-*Rhizobium* Symbiosis. Annual Review of *Plant Biology 68*, 535-561.

Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. *Bioinformatics 26*, 2363-2367.

Castaings, L., Camargo, A., Pocholle, D., Gaudon, V., Texier, Y., Boutet-Mercey, S., Taconnat, L., Renou, J.P., Daniel-Vedele, F., Fernandez, E.*, et al.* (2009). The nodule inception-like protein 7 modulates nitrate sensing and metabolism in *Arabidopsis*. *Plant J 57*, 426-435.

Cheng, X., Wang, M., Lee, H.K., Tadege, M., Ratet, P., Udvardi, M., Mysore, K.S., and Wen, J. (2014). An efficient reverse genetics platform in the model legume *Medicago truncatula*. *New Phytol 201*, 1065-1076.

Chiasson, D.M., Loughlin, P.C., Mazurkiewicz, D., Mohammadidehcheshmeh, M., Fedorova, E.E., Okamoto, M., McLean, E., Glass, A.D., Smith, S.E., Bisseling, T.*, et al.* (2014). Soybean *SAT1* (*Symbiotic Ammonium Transporter 1*) encodes a bHLH transcription factor involved in nodule growth and NH4+ transport. *Proc Natl Acad Sci* U S A *111*, 4814-4819.

Cooper, J.E. (2007). Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue. *J Appl Microbiol 103*, 1355-1365.

Covington, M.F., Maloof, J.N., Straume, M., Kay, S.A., and Harmer, S.L. (2008). Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biol 9*, R130.

Czernic, P., Gully, D., Cartieaux, F., Moulin, L., Guefrachi, I., Patrel, D., Pierre, O., Fardoux, J., Chaintreuil, C., Nguyen, P.*, et al.* (2015). Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich Peptides. *Plant Physiol 169*, 1254-1265.

d'Erfurth, I., Cosson, V., Eschstruth, A., Lucas, H., Kondorosi, A., and Ratet, P. (2003). Efficient transposition of the Tnt1 tobacco retrotransposon in the model legume *Medicago truncatula*. *Plant J 34*, 95-106.

Dalla Via, V., Narduzzi, C., Aguilar, O.M., Zanetti, M.E., and Blanco, F.A. (2015). Changes in the Common Bean Transcriptome in Response to Secreted and Surface Signal Molecules of *Rhizobium etli*. *Plant Physiol 169*, 1356-1370.

De Cuyper, C., Fromentin, J., Yocgo, R.E., De Keyser, A., Guillotin, B., Kunert, K., Boyer, F.D., and Goormachtig, S. (2015). From lateral root density to nodule number, the strigolactone analogue GR24 shapes the root architecture of *Medicago truncatula*. *J Exp Bot 66*, 4091.

de Montaigu, A., Toth, R., and Coupland, G. (2010). Plant development goes like clockwork. *Trends Genet 26*, 296-306.

De Smet, I., Tetsumura, T., De Rybel, B., Frei dit Frey, N., Laplaze, L., Casimiro, I., Swarup, R., Naudts, M., Vanneste, S., Audenaert, D.*, et al.* (2007). Auxin-dependent regulation of lateral root positioning in the basal meristem of *Arabidopsis*. *Development 134*, 681-690.

Ding, Y., Kalo, P., Yendrek, C., Sun, J., Liang, Y., Marsh, J.F., Harris, J.M., and Oldroyd, G.E. (2008). Abscisic acid coordinates nod factor and cytokinin signaling during the regulation of nodulation in *Medicago truncatula*. *Plant Cell 20*, 2681-2695.

Dodd, A.N., Salathia, N., Hall, A., Kévei, E., Tóth, R., Nagy, F., Hibberd, J.M., Millar, A.J., and Webb, A.A.R. (2005). Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage. *Science 309*, 630-633.

Durgo, H., Klement, E., Hunyadi-Gulyas, E., Szucs, A., Kereszt, A., Medzihradszky, K.F., and Kondorosi, E. (2015). Identification of nodule-specific cysteine-rich plant peptides in endosymbiotic bacteria. *Proteomics 15*, 2291-2295.

Edwards, K.D., Lynn, J.R., Gyula, P., Nagy, F., and Millar, A.J. (2005). Natural allelic variation in the temperature-compensation mechanisms of the *Arabidopsis thaliana* circadian clock. *Genetics 170*, 387-400.

Farkas, A., Maroti, G., Durgo, H., Gyorgypal, Z., Lima, R.M., Medzihradszky, K.F., Kereszt, A., Mergaert, P., and Kondorosi, E. (2014). *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proc Natl Acad Sci* U S A *111*, 5183-5188.

Farre, E.M. (2012). The regulation of plant growth by the circadian clock. Plant Biol (Stuttg) *14*, 401-410.

Fedorova, M., van de Mortel, J., Matsumoto, P.A., Cho, J., Town, C.D., VandenBosch, K.A., Gantt, J.S., and Vance, C.P. (2002). Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiol 130*, 519-537.

Ferguson, B.J., Indrasumunar, A., Hayashi, S., Lin, M.H., Lin, Y.H., Reid, D.E., and Gresshoff, P.M. (2010). Molecular analysis of legume nodule development and autoregulation. *J Integr Plant Biol 52*, 61-76.

Ferguson, B.J., Ross, J.J., and Reid, J.B. (2005). Nodulation Phenotypes of Gibberellin and Brassinosteroid Mutants of Pea. *Plant Physiology 138*, 2396-2405.

Flagel, L.E., and Wendel, J.F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol 183*, 557-564.

Franco-Zorrilla, J.M., Lopez-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci* U S A *111*, 2367-2372.

Frugier, F., Kosuta, S., Murray, J.D., Crespi, M., and Szczyglowski, K. (2008). Cytokinin: secret agent of symbiosis. *Trends Plant Sci 13*, 115-120.

Fu, Y.H., and Marzluf, G.A. (1990). nit-2, the major nitrogen regulatory gene of Neurospora crassa, encodes a protein with a putative zinc finger DNA-binding domain. *Mol Cell Biol 10*, 1056-1065.

Gansel, X., Munos, S., Tillard, P., and Gojon, A. (2001). Differential regulation of the NO3- and NH4+ transporter genes *AtNrt2.1* and *AtAmt1.1* in *Arabidopsis*: relation with long-distance and local controls by N status of the plant. Plant J *26*, 143-155.

Gazzarrini, S., Lejay, L., Gojon, A., Ninnemann, O., Frommer, W.B., and von Wiren, N. (1999). Three functional transporters for constitutive, diurnally regulated, and starvation-induced uptake of ammonium into *Arabidopsis* roots. *Plant Cell 11*, 937-948.

Gibson, K.E., Kobayashi, H., and Walker, G.C. (2008). Molecular determinants of a symbiotic chronic infection. *Annu Rev Genet 42*, 413-441.

Gifford, M.L., Dean, A., Gutierrez, R.A., Coruzzi, G.M., and Birnbaum, K.D. (2008). Cell-specific nitrogen responses mediate developmental plasticity. *Proc Natl Acad Sci* U S A *105*, 803-808.

Giordano, W., and Hirsch, A.M. (2004). The expression of MaEXP1, a Melilotus alba expansin gene, is upregulated during the sweetclover-*Sinorhizobium meliloti* interaction. *Mol Plant Microbe Interact 17*, 613-622.

Goodspeed, D., Chehab, E.W., Min-Venditti, A., Braam, J., and Covington, M.F. (2012). *Arabidopsis* synchronizes jasmonate-mediated defense with insect circadian behavior. *Proc Natl Acad Sci* U S A *109*, 4674-4677.

Graham, M.A., Silverstein, K.A., Cannon, S.B., and VandenBosch, K.A. (2004). Computational identification and characterization of novel genes from legumes. *Plant Physiol 135*, 1179-1197.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics 27*, 1017-1018.

Green, R.M., Tingay, S., Wang, Z.-Y., and Tobin, E.M. (2002). Circadian Rhythms Confer a Higher Level of Fitness to *Arabidopsis* Plants. *Plant Physiology 129*, 576-584.

Gresshoff, P.M., Lohar, D., Chan, P.K., Biswas, B., Jiang, Q., Reid, D., Ferguson, B., and Stacey, G. (2009). Genetic analysis of ethylene regulation of legume nodulation. *Plant Signal Behav 4*, 818-823.

Grunewald, W., van Noorden, G., Van Isterdael, G., Beeckman, T., Gheysen, G., and Mathesius, U. (2009). Manipulation of auxin transport in plant roots during *Rhizobium* symbiosis and nematode parasitism. *Plant Cell 21*, 2553-2562.

Guefrachi, I., Nagymihaly, M., Pislariu, C.I., Van de Velde, W., Ratet, P., Mars, M., Udvardi, M.K., Kondorosi, E., Mergaert, P., and Alunni, B. (2014). Extreme specificity of *NCR* gene expression in *Medicago truncatula*. BMC Genomics *15*, 712.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. *Genome Biol 8*, R24.

Gutierrez, R.A. (2012). Systems biology for enhanced plant nitrogen nutrition. *Science 336*, 1673-1675.

Haag, A.F., Arnold, M.F., Myka, K.K., Kerscher, B., Dall'Angelo, S., Zanda, M., Mergaert, P., and Ferguson, G.P. (2013). Molecular insights into bacteroid development during *Rhizobium*-legume symbiosis. *FEMS Microbiol Rev 37*, 364-383.

Harmer, S.L. (2009). The circadian system in higher plants. *Annu Rev Plant Biol 60*, 357-377.

Harmer, S.L., Hogenesch, J.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A., and Kay, S.A. (2000). Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science 290*, 2110-2113.

Harmer, S.L., and Kay, S.A. (2005). Positive and negative factors confer phase-specific circadian regulation of transcription in *Arabidopsis*. *Plant Cell 17*, 1926-1940.

Haydon, M.J., Mielczarek, O., Robertson, F.C., Hubbard, K.E., and Webb, A.A.R. (2013). Photosynthetic entrainment of the *Arabidopsis thaliana* circadian clock. *Nature 502*, 689.

Haydon, M.J., Roman, A., and Arshad, W. (2015). Nutrient homeostasis within the plant circadian network. *Front Plant Sci 6*, 299.

Heath, K.D., and Tiffin, P. (2009). Stabilizing mechanisms in a legume-*rhizobium* mutualism. *Evolution 63*, 652-662.

Hecht, V., Foucher, F., Ferrandiz, C., Macknight, R., Navarro, C., Morin, J., Vardy, M.E., Ellis, N., Beltran, J.P., Rameau, C*., et al.* (2005). Conservation of *Arabidopsis* flowering genes in model legumes. *Plant Physiol 137*, 1420-1434.

Hecht, V., Knowles, C.L., Vander Schoor, J.K., Liew, L.C., Jones, S.E., Lambert, M.J., and Weller, J.L. (2007). *Pea LATE BLOOMER1* is a *GIGANTEA* ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. *Plant Physiol 144*, 648-661.

Heckmann, A.B., Sandal, N., Bek, A.S., Madsen, L.H., Jurkiewicz, A., Nielsen, M.W., Tirichine, L., and Stougaard, J. (2011). Cytokinin induction of root nodule primordia in *Lotus japonicus* is regulated by a mechanism operating in the root cortex. *Mol Plant Microbe Interact 24*, 1385-1395.

Hevia, M.A., Canessa, P., Muller-Esparza, H., and Larrondo, L.F. (2015). A circadian oscillator in the fungus *Botrytis cinerea* regulates virulence when infecting *Arabidopsis thaliana*. *Proc Natl Acad Sci* U S A *112*, 8744-8749.

Hirel, B., and Lea, P.J. (2001). Ammonia Assimilation. In Plant Nitrogen, P.J. Lea, and J.-F. Morot-Gaudry, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 79-99.

Hirel, B., Tétu, T., Lea, P.J., and Dubois, F. (2011). Improving Nitrogen Use Efficiency in Crops for Sustainable Agriculture. *Sustainability 3*, 1452.

Hirsch, A.M., Bhuvaneswari, T.V., Torrey, J.G., and Bisseling, T. (1989). Early nodulin genes are induced in alfalfa root outgrowths elicited by auxin transport inhibitors. *Proc Natl Acad Sci* U S A *86*, 1244-1248.

Hirsch, A.M., Larue, T.A., and Doyle, J. (1997). Is the Legume Nodule a Modified Root or Stem or an Organ sui generis? *Critical Reviews in Plant Sciences 16*, 361-392.

Horvath, B., Domonkos, A., Kereszt, A., Szucs, A., Abraham, E., Ayaydin, F., Boka, K., Chen, Y., Chen, R., Murray, J.D.*, et al.* (2015). Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the Medicago truncatula dnf7 mutant. *Proc Natl Acad Sci* U S A *112*, 15232-15237.

Huault, E., Laffont, C., Wen, J., Mysore, K.S., Ratet, P., Duc, G., and Frugier, F. (2014). Local and systemic regulation of plant root system architecture and symbiotic nodulation by a receptor-like kinase. *PLoS Genet 10*, e1004891.

HwangBo, K., Son, S.H., Lee, J.S., Min, S.R., Ko, S.M., Liu, J.R., Choi, D., and Jeong, W.J. (2010). Rapid and simple method for DNA extraction from plant and algal species suitable for PCR amplification using a chelating resin Chelex 100. *Plant Biotechnology Reports 4*, 49-52.

Iannetta, P.P., Young, M., Bachinger, J., Bergkvist, G., Doltra, J., Lopez-Bellido, R.J., Monti, M., Pappa, V.A., Reckling, M., Topp, C.F.*, et al.* (2016). A Comparative Nitrogen Balance and Productivity Analysis of Legume and Non-legume Supported Cropping Systems: The Potential Role of Biological Nitrogen Fixation. *Front Plant Sci 7*, 1700.

Iijima, M., and Matsushita, N. (2011). A circadian and an ultradian rhythm are both evident in root growth of rice. *J Plant Physiol 168*, 2072-2080.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics 4*, 249-264.

Ishikawa, K., Yokota, K., Li, Y.Y., Wang, Y., Liu, C.-T., Suzuki, S., Aono, T., and Oyaizu, H. (2008). Isolation of a novel root-determined hypernodulation mutant *rdh1* of *Lotus japonicus*. *Soil Science and Plant Nutrition 54*, 259-263.

Ivanchenko, M.G., Muday, G.K., and Dubrovsky, J.G. (2008). Ethylene-auxin interactions regulate lateral root initiation and emergence in *Arabidopsis thaliana*. *Plant J 55*, 335-347.

James, A.B., Monreal, J.A., Nimmo, G.A., Kelly, C.L., Herzyk, P., Jenkins, G.I., and Nimmo, H.G. (2008). The circadian clock in *Arabidopsis* roots is a simplified slave version of the clock in shoots. *Science 322*, 1832-1835.

Jeong, M.J., and Shih, M.C. (2003). Interaction of a GATA factor with cis-acting elements involved in light regulation of nuclear genes encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase in *Arabidopsis*. *Biochem Biophys Res Commun 300*, 555-562.

Jeong, S., Trotochaud, A.E., and Clark, S.E. (1999). The *Arabidopsis* CLAVATA2 gene encodes a receptor-like protein required for the stability of the CLAVATA1 receptor-like kinase. *Plant Cell 11*, 1925-1934.

Jeudy, C., Ruffel, S., Freixes, S., Tillard, P., Santoni, A.L., Morel, S., Journet, E.P., Duc, G., Gojon, A., Lepetit, M.*, et al.* (2010). Adaptation of *Medicago truncatula* to nitrogen limitation is modulated via local and systemic nodule developmental responses. *New Phytol 185*, 817-828.

Jin, J., Watt, M., and Mathesius, U. (2012). The autoregulation gene *SUNN* mediates changes in root organ formation in response to nitrogen through alteration of shoot-to-root auxin transport. *Plant Physiol 159*, 489-500.

Jones, K.M., Kobayashi, H., Davies, B.W., Taga, M.E., and Walker, G.C. (2007). How rhizobial symbionts invade plants: the *Sinorhizobium–Medicago* model. *Nature Reviews Microbiology 5*, 619.

Kakar, K., Wandrey, M., Czechowski, T., Gaertner, T., Scheible, W.R., Stitt, M., Torres-Jerez, I., Xiao, Y., Redman, J.C., Wu, H.C.*, et al.* (2008). A community resource for high-throughput quantitative RT-PCR analysis of transcription factor gene expression in *Medicago truncatula*. Plant Methods *4*, 18.

Kalo, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R.M., Hirsch, S., Jakab, J., Sims, S., Long, S.R., Rogers, J.*, et al.* (2005). Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science 308*, 1786-1789.

Kassaw, T., Bridges, W., Jr., and Frugoli, J. (2015). Multiple Autoregulation of Nodulation (AON) Signals Identified through Split Root Analysis of *Medicago truncatula sunn* and *rdn1* Mutants. *Plants* (Basel) *4*, 209-224.

Kato, T., Kawashima, K., Miwa, M., Mimura, Y., Tamaoki, M., Kouchi, H., and Suganuma, N. (2002). Expression of genes encoding late nodulins characterized by a putative signal peptide and conserved cysteine residues is reduced in ineffective pea nodules. *Mol Plant Microbe Interact 15*, 129-137.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res 30*, 3059-3066.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol 30*, 772-780.

Kereszt, A., Mergaert, P., and Kondorosi, E. (2011). Bacteroid development in legume nodules: evolution of mutual benefit or of sacrificial victims? *Mol Plant Microbe Interact 24*, 1300-1309.

Kim, J., Jung, J.H., Reyes, J.L., Kim, Y.S., Kim, S.Y., Chung, K.S., Kim, J.A., Lee, M., Lee, Y., Narry Kim, V.*, et al.* (2005). microRNA-directed cleavage of *ATHB15* mRNA regulates vascular development in *Arabidopsis* inflorescence stems. *Plant J 42*, 84-94.

Kim, M., Chen, Y., Xi, J., Waters, C., Chen, R., and Wang, D. (2015). An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proc Natl Acad Sci* U S A *112*, 15238-15243.

Kondorosi, E., Mergaert, P., and Kereszt, A. (2013). A paradigm for endosymbiotic life: cell differentiation of *Rhizobium* bacteria provoked by host plant factors. *Annu Rev Microbiol 67*, 611-628.

Kouchi, H., Imaizumi-Anraku, H., Hayashi, M., Hakoyama, T., Nakagawa, T., Umehara, Y., Suganuma, N., and Kawaguchi, M. (2010). How many peas in a pod? Legume genes responsible for mutualistic symbioses underground. *Plant Cell Physiol 51*, 1381-1397.

Kraiser, T., Gras, D.E., Gutierrez, A.G., Gonzalez, B., and Gutierrez, R.A. (2011). A holistic view of nitrogen acquisition in plants. *J Exp Bot 62*, 1455-1466.

Krapp, A., David, L.C., Chardin, C., Girin, T., Marmagne, A., Leprince, A.S., Chaillou, S., Ferrario-Mery, S., Meyer, C., and Daniel-Vedele, F. (2014). Nitrate transport and signalling in *Arabidopsis*. *J Exp Bot 65*, 789-798.

Krusell, L., Madsen, L.H., Sato, S., Aubert, G., Genua, A., Szczyglowski, K., Duc, G., Kaneko, T., Tabata, S., de Bruijn, F*., et al.* (2002). Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature 420*, 422-426.

Kuppusamy, K.T., Ivashuta, S., Bucciarelli, B., Vance, C.P., Gantt, J.S., and Vandenbosch, K.A. (2009). Knockdown of CELL DIVISION CYCLE16 reveals an inverse relationship between lateral root and nodule numbers and a link to auxin in *Medicago truncatula*. *Plant Physiol 151*, 1155-1166.

Lam, H.M., Coschigano, K.T., Oliveira, I.C., Melo-Oliveira, R., and Coruzzi, G.M. (1996). The Molecular-Genetics of Nitrogen Assimilation into Amino Acids in Higher Plants. *Annu Rev Plant Physiol Plant Mol Biol 47*, 569-593.

Larrainzar, E., Riely, B.K., Kim, S.C., Carrasquilla-Garcia, N., Yu, H.J., Hwang, H.J., Oh, M., Kim, G.B., Surendrarao, A.K., Chasman, D*., et al.* (2015). Deep Sequencing of the *Medicago truncatula* Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals. *Plant Physiol 169*, 233-265.

Lavin, M., Doyle, J.J., and Palmer, J.D. (1990). Evolutionary Significance of the Loss of the Chloroplast-DNA Inverted Repeat in the Leguminosae Subfamily Papilionoideae. *Evolution 44*, 390-402.

Lavin, M., Herendeen, P.S., and Wojciechowski, M.F. (2005). Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Syst Biol 54*, 575-594.

Lee, D.K., Ahn, J.H., Song, S.K., Choi, Y.D., and Lee, J.S. (2003). Expression of an expansin gene is correlated with root elongation in *soybean*. *Plant Physiol 131*, 985-997.

Lejay, L., Tillard, P., Lepetit, M., Olive, F., Filleur, S., Daniel-Vedele, F., and Gojon, A. (1999). Molecular and functional regulation of two NO3- uptake systems by N- and C-status of *Arabidopsis* plants. Plant J *18*, 509-519.

Leran, S., Varala, K., Boyer, J.C., Chiurazzi, M., Crawford, N., Daniel-Vedele, F., David, L., Dickstein, R., Fernandez, E., Forde, B.*, et al.* (2014). A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. *Trends Plant Sci 19*, 5-9.

Lewis, D.R., Negi, S., Sukumar, P., and Muday, G.K. (2011). Ethylene inhibits lateral root development, increases IAA transport and expression of *PIN3* and *PIN7* auxin efflux carriers. *Development 138*, 3485-3495.

Li, D., Kinkema, M., and Gresshoff, P.M. (2009a). Autoregulation of nodulation (AON) in *Pisum sativum* (pea) involves signalling events associated with both nodule primordia development and nitrogen fixation. *J Plant Physiol 166*, 955-967.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009b). The Sequence Alignment/Map format and SAMtools. *Bioinformatics 25*, 2078-2079.

Liew, L.C., Hecht, V., Laurie, R.E., Knowles, C.L., Vander Schoor, J.K., Macknight, R.C., and Weller, J.L. (2009). *DIE NEUTRALIS* and *LATE BLOOMER 1* contribute to regulation of the pea circadian clock. *Plant Cell 21*, 3198-3211.

Liu, H., Zhang, C., Yang, J., Yu, N., and Wang, E. (2018). Hormone modulation of legume-rhizobial symbiosis. *Journal of Integrative Plant Biology 60*, 632-648.

Lohar, D., Stiller, J., Kam, J., Stacey, G., and Gresshoff, P.M. (2009). Ethylene insensitivity conferred by a mutated *Arabidopsis* ethylene receptor gene alters nodulation in transgenic *Lotus japonicus*. *Ann Bot 104*, 277-285.

Lopez-Bucio, J., Cruz-Ramirez, A., and Herrera-Estrella, L. (2003). The role of nutrient availability in regulating root architecture. *Curr Opin Plant Biol 6*, 280-287.

López-Juez, E., and Devlin, P.F. (2008). Light and the Control of Plant Growth. In Plant Growth Signaling, L. Bögre, and G. Beemster, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 223-242.

Lu, H., Zou, Y., and Feng, N. (2010). Overexpression of AHL20 negatively regulates defenses in *Arabidopsis*. *J Integr Plant Biol 52*, 801-808.

Madsen, L.H., Tirichine, L., Jurkiewicz, A., Sullivan, J.T., Heckmann, A.B., Bek, A.S., Ronson, C.W., James, E.K., and Stougaard, J. (2010). The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nat Commun 1*, 10.

Magori, S., Oka-Kira, E., Shibata, S., Umehara, Y., Kouchi, H., Hase, Y., Tanaka, A., Sato, S., Tabata, S., and Kawaguchi, M. (2009). Too much love, a root regulator associated with the long-distance control of nodulation in *Lotus japonicus*. *Mol Plant Microbe Interact 22*, 259-268.

Maher, K.A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D.A., Zumstein, K., Woodhouse, M., Bubb, K., Dorrity, M.W.*, et al.* (2018). Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules. *Plant Cell 30*, 15-36.

Maillet, F., Poinsot, V., Andre, O., Puech-Pages, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A.*, et al.* (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature 469*, 58-63.

Manfield, I.W., Devlin, P.F., Jen, C.H., Westhead, D.R., and Gilmartin, P.M. (2007). Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the *Arabidopsis* GATA gene family. *Plant Physiol 143*, 941-958.

Marchive, C., Roudier, F., Castaings, L., Brehaut, V., Blondet, E., Colot, V., Meyer, C., and Krapp, A. (2013). Nuclear retention of the transcription factor NLP7 orchestrates the early response to nitrate in plants. *Nat Commun 4*, 1713.

Marcolino-Gomes, J., Rodrigues, F.A., Fuganti-Pagliarini, R., Bendix, C., Nakayama, T.J., Celaya, B., Molinari, H.B., de Oliveira, M.C., Harmon, F.G., and Nepomuceno, A. (2014). Diurnal oscillations of soybean circadian clock and drought responsive genes. *PLoS One 9*, e86402.

Maroti, G., Kereszt, A., Kondorosi, E., and Mergaert, P. (2011). Natural roles of antimicrobial peptides in microbes, plants and animals. *Res Microbiol 162*, 363-374.

Maroti, G., and Kondorosi, E. (2014). Nitrogen-fixing Rhizobium-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Front Microbiol 5*, 326.

Marshall, E., Costa, L.M., and Gutierrez-Marcos, J. (2011). Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J Exp Bot 62*, 1677-1686.

Matamoros, M.A., Baird, L.M., Escuredo, P.R., Dalton, D.A., Minchin, F.R., Iturbe-Ormaetxe, I., Rubio, M.C., Moran, J.F., Gordon, A.J., and Becana, M. (1999). Stress-Induced Legume Root Nodule Senescence. Physiological, Biochemical, and Structural Alterations. *Plant Physiology 121*, 97-112.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R.*, et al.* (2016). JASPAR 2016: a major expansion

and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res 44*, D110-115.

Mathesius, U. (2003). Conservation and divergence of signalling pathways between roots and soil microbes – the Rhizobium-legume symbiosis compared to the development of lateral roots, mycorrhizal interactions and nematode-induced galls. *Plant and Soil 255*, 105-119.

Mathesius, U., Schlaman, H.R., Spaink, H.P., Of Sautter, C., Rolfe, B.G., and Djordjevic, M.A. (1998). Auxin transport inhibition precedes root nodule formation in white clover roots and is regulated by flavonoids and derivatives of chitin oligosaccharides. *Plant J 14*, 23-34.

Matsushita, A., Furumoto, T., Ishida, S., and Takahashi, Y. (2007). AGF1, an AT-hook protein, is necessary for the negative feedback of *AtGA3ox1* encoding GA 3-oxidase. *Plant Physiol 143*, 1152-1162.

McClung, C.R., and Gutierrez, R.A. (2010). Network news: prime time for systems biology of the plant circadian clock. *Curr Opin Gene*t Dev *20*, 588-598.

Mergaert, P., Kikuchi, Y., Shigenobu, S., and Nowack, E.C.M. (2017). Metabolic Integration of Bacterial Endosymbionts through Antimicrobial Peptides. *Trends Microbiol 25*, 703-712.

Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., and Kondorosi, E. (2003). A novel family in Medicago truncatula consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiol 132*, 161-173.

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A.E., Barloy-Hubler, F., Galibert, F., Kondorosi, A*., et al.* (2006). Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc Natl Acad Sci* U S A *103*, 5230-5235.

Michael, T.P., and McClung, C.R. (2003). Enhancer trapping reveals widespread circadian clock transcriptional control in *Arabidopsis*. *Plant Physiol 132*, 629-639.

Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer, A., Trout, J.D., Hazen, S.P., Shen, R., Priest, H.D., Sullivan, C.M.*, et al.* (2008). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet 4*, e14.

Mikulass, K.R., Nagy, K., Bogos, B., Szegletes, Z., Kovacs, E., Farkas, A., Varo, G., Kondorosi, E., and Kereszt, A. (2016). Antimicrobial nodule-specific cysteine-rich peptides disturb the integrity of bacterial outer and inner membranes and cause loss of membrane potential. *Ann Clin Microbiol Antimicrob 15*, 43.

Miller, A.J., Fan, X., Orsel, M., Smith, S.J., and Wells, D.M. (2007). Nitrate transport and signalling. *J Exp Bot 58*, 2297-2306.

Minchin, F.R., and Pate, J.S. (1974). Diurnal Functioning of the Legume Root Nodule. *Journal of Experimental Botany 25*, 295-308.

Miyazawa, H., Oka-Kira, E., Sato, N., Takahashi, H., Wu, G.J., Sato, S., Hayashi, M., Betsuyaku, S., Nakazono, M., Tabata, S.*, et al.* (2010). The receptor-like kinase KLAVIER mediates systemic regulation of nodulation and non-symbiotic shoot development in Lotus japonicus. *Development 137*, 4317-4325.

Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H.R., Carre, I.A., and Coupland, G. (2002). *LHY* and *CCA1* are partially redundant genes required to maintain circadian rhythms in *Arabidopsis*. *Dev Cell 2*, 629-641.

Morere-Le Paven, M.C., Viau, L., Hamon, A., Vandecasteele, C., Pellizzaro, A., Bourdin, C., Laffont, C., Lapied, B., Lepetit, M., Frugier, F.*, et al.* (2011). Characterization of a dual-affinity nitrate transporter *MtNRT1.3* in the model legume *Medicago truncatula*. *J Exp Bot 62*, 5595-5605.

Morgulis, A., Gertz, E.M., Schaffer, A.A., and Agarwala, R. (2006). WindowMasker: window-based masker for sequenced genomes. *Bioinformatics 22*, 134-141.

Mortier, V., De Wever, E., Vuylsteke, M., Holsters, M., and Goormachtig, S. (2012a). Nodule numbers are governed by interaction between CLE peptides and cytokinin signaling. *Plant J 70*, 367-376.

Mortier, V., Holsters, M., and Goormachtig, S. (2012b). Never too many? How legumes control nodule numbers. *Plant Cell Environ 35*, 245-258.

Mounier, E., Pervent, M., Ljung, K., Gojon, A., and Nacry, P. (2014). Auxin-mediated nitrate signalling by NRT1.1 participates in the adaptive response of *Arabidopsis* root architecture to the spatial heterogeneity of nitrate availability. *Plant Cell Environ 37*, 162-174.

Mount, D.W. (2008). Using gaps and gap penalties to optimize pairwise sequence alignments. *CSH Protoc*, pdb top40.

Muller, R., Bleckmann, A., and Simon, R. (2008). The receptor kinase CORYNE of *Arabidopsis* transmits the stem cell-limiting signal CLAVATA3 independently of CLAVATA1. *Plant Cell 20*, 934-946.

Nagel, D.H., Doherty, C.J., Pruneda-Paz, J.L., Schmitz, R.J., Ecker, J.R., and Kay, S.A. (2015). Genome-wide identification of *CCA1* targets uncovers an expanded clock network in *Arabidopsis*. *Proc Natl Acad Sci* U S A *112*, E4802-4810.

Nakahata, Y., Grimaldi, B., Sahar, S., Hirayama, J., and Sassone-Corsi, P. (2007). Signaling to the circadian clock: plasticity by chromatin remodeling. *Curr Opin Cell Biol 19*, 230-237.

Nallu, S., Silverstein, K.A., Samac, D.A., Bucciarelli, B., Vance, C.P., and VandenBosch, K.A. (2013). Regulatory patterns of a large family of defensin-like genes expressed in nodules of *Medicago truncatula*. *PLoS One 8*, e60355.

Nallu, S., Silverstein, K.A., Zhou, P., Young, N.D., and Vandenbosch, K.A. (2014). Patterns of divergence of a large family of nodule cysteine-rich peptides in accessions of *Medicago truncatula*. *Plant J 78*, 697-705.

Navarro, L., Bari, R., Achard, P., Lison, P., Nemri, A., Harberd, N.P., and Jones, J.D. (2008). DELLAs control plant immune responses by modulating the balance of jasmonic acid and salicylic acid signaling. *Curr Biol 18*, 650-655.

Nishida, H., Tanaka, S., Handa, Y., Ito, M., Sakamoto, Y., Matsunaga, S., Betsuyaku, S., Miura, K., Soyano, T., Kawaguchi, M*., et al.* (2018). A NIN-LIKE PROTEIN mediates nitrate-induced control of root nodule symbiosis in *Lotus japonicus*. *Nat Commun 9*, 499.

Nishimura, R., Ohmori, M., Fujita, H., and Kawaguchi, M. (2002). A Lotus basic leucine zipper protein with a RING-finger motif negatively regulates the developmental program of nodulation. *Proc Natl Acad Sci* U S A *99*, 15206-15210.

Nozue, K., Covington, M.F., Duek, P.D., Lorrain, S., Fankhauser, C., Harmer, S.L., and Maloof, J.N. (2007). Rhythmic growth explained by coincidence between internal and external cues. *Nature 448*, 358-361.

O'Brien, J.A., Vega, A., Bouguyon, E., Krouk, G., Gojon, A., Coruzzi, G., and Gutierrez, R.A. (2016). Nitrate Transport, Sensing, and Responses in Plants. *Mol Plant 9*, 837-856.

Ochando, I., Jover-Gil, S., Ripoll, J.J., Candela, H., Vera, A., Ponce, M.R., Martinez-Laborda, A., and Micol, J.L. (2006). Mutations in the microRNA complementarity site of the *INCURVATA4* gene perturb meristem function and adaxialize lateral organs in arabidopsis. *Plant Physiol 141*, 607-619.

Ohashi-Ito, K., and Fukuda, H. (2003). *HD-zip III* homeobox genes that include a novel member, *ZeHB-13 (Zinnia)/ATHB-15* (Arabidopsis), are involved in procambium and xylem cell differentiation. *Plant Cell Physiol 44*, 1350-1358.

Oka-Kira, E., and Kawaguchi, M. (2006). Long-distance signaling to control root nodule number. *Curr Opin Plant Biol 9*, 496-502.

Okamoto, S., Ohnishi, E., Sato, S., Takahashi, H., Nakazono, M., Tabata, S., and Kawaguchi, M. (2009). Nod factor/nitrate-induced *CLE* genes that drive HAR1-mediated systemic regulation of nodulation. *Plant Cell Physiol 50*, 67-77.

Olah, B., Briere, C., Becard, G., Denarie, J., and Gough, C. (2005). Nod factors and a diffusible factor from arbuscular mycorrhizal fungi stimulate lateral root formation in *Medicago truncatula* via the DMI1/DMI2 signalling pathway. Plant J *44*, 195-207.

Oldroyd, G.E. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat Rev Microbiol 11*, 252-263.

Oldroyd, G.E., Engstrom, E.M., and Long, S.R. (2001). Ethylene inhibits the Nod factor signal transduction pathway of *Medicago truncatula*. *Plant Cell 13*, 1835-1849.

Oldroyd, G.E., Murray, J.D., Poole, P.S., and Downie, J.A. (2011). The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet 45*, 119-144.

Oliveira, I.C., and Coruzzi, G.M. (1999). Carbon and amino acids reciprocally modulate the expression of glutamine synthetase in *Arabidopsis*. Plant Physiol *121*, 301-310.

Oono, R., Anderson, C.G., and Denison, R.F. (2011). Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc Biol Sci 278*, 2698-2703.

Oono, R., and Denison, R.F. (2010). Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiol 154*, 1541-1548.

Oono, R., Schmitt, I., Sprent, J.I., and Denison, R.F. (2010). Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. *New Phytol 187*, 508-520.

Ordogh, L., Voros, A., Nagy, I., Kondorosi, E., and Kereszt, A. (2014). Symbiotic plant peptides eliminate *Candida albicans* both in vitro and in an epithelial infection model and inhibit the proliferation of immortalized human cells. *Biomed Res Int 2014*, 320796.

Pan, H., and Wang, D. (2017). Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. *Nat Plants 3*, 17048.

Patterson, K., Cakmak, T., Cooper, A., Lager, I., Rasmusson, A.G., and Escobar, M.A. (2010). Distinct signalling pathways and transcriptome response signatures differentiate ammonium- and nitrate-supplied plants. *Plant Cell Environ 33*, 1486-1501.

Penmetsa, R.V., Frugoli, J.A., Smith, L.S., Long, S.R., and Cook, D.R. (2003). Dual genetic pathways controlling nodule number in *Medicago truncatula*. Plant Physiol *131*, 998-1008.

Penmetsa, R.V., Uribe, P., Anderson, J., Lichtenzveig, J., Gish, J.C., Nam, Y.W., Engstrom, E., Xu, K., Sckisel, G., Pereira, M.*, et al.* (2008). The *Medicago truncatula* ortholog of Arabidopsis EIN2, sickle, is a negative regulator of symbiotic and pathogenic microbial associations. *Plant J 55*, 580-595.

Penterman, J., Abo, R.P., De Nisco, N.J., Arnold, M.F., Longhi, R., Zanda, M., and Walker, G.C. (2014). Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. *Proc Natl Acad Sci* U S A *111*, 3561-3566.

Peret, B., De Rybel, B., Casimiro, I., Benkova, E., Swarup, R., Laplaze, L., Beeckman, T., and Bennett, M.J. (2009). *Arabidopsis* lateral root development: an emerging story. *Trends Plant Sci 14*, 399-408.
Perret, X., Staehelin, C., and Broughton, W.J. (2000). Molecular basis of symbiotic promiscuity. *Microbiol Mol Biol Rev 64*, 180-201.

Peters, N.K., and Crist-Estes, D.K. (1989). Nodule formation is stimulated by the ethylene inhibitor aminoethoxyvinylglycine. *Plant Physiol 91*, 690-693.

Picard, K., Lee, R., Hellens, R., and Macknight, R. (2013). Transient gene expression in *Medicago truncatula* leaves via Agroinfiltration. Methods Mol Biol *1069*, 215-226.

Pruneda-Paz, J.L., and Kay, S.A. (2010). An expanding universe of circadian networks in higher plants. *Trends Plant Sci 15*, 259-265.

Rastogi, R., Bate, N.J., Sivasankar, S., and Rothstein, S.J. (1997). Footprinting of the spinach nitrite reductase gene promoter reveals the preservation of nitrate regulatory elements between fungi and higher plants. *Plant Mol Biol 34*, 465-476.

Raven, J.A., and Edwards, D. (2001). Roots: evolutionary origins and biogeochemical significance. *J Exp Bot 52*, 381-401.

Rawat, R., Schwartz, J., Jones, M.A., Sairanen, I., Cheng, Y., Andersson, C.R., Zhao, Y., Ljung, K., and Harmer, S.L. (2009). REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. *Proc Natl Acad Sci* U S A *106*, 16883-16888.

Reid, D.E., Ferguson, B.J., Hayashi, S., Lin, Y.H., and Gresshoff, P.M. (2011). Molecular mechanisms controlling legume autoregulation of nodulation. *Ann Bot 108*, 789-795.

Reid, D.E., Heckmann, A.B., Novak, O., Kelly, S., and Stougaard, J. (2016). CYTOKININ OXIDASE/DEHYDROGENASE3 Maintains Cytokinin Homeostasis during Root and Nodule Development in *Lotus japonicus*. *Plant Physiol 170*, 1060-1074.

Reyes, J.C., Muro-Pastor, M.I., and Florencio, F.J. (2004). The GATA family of transcription factors in Arabidopsis and rice. Plant Physiol *134*, 1718-1732.

Ripperger, J.A., and Merrow, M. (2011). Perfect timing: epigenetic regulation of the circadian clock. *FEBS Lett 585*, 1406-1411.

Roux, B., Rodde, N., Jardinaud, M.F., Timmers, T., Sauviac, L., Cottret, L., Carrere, S., Sallet, E., Courcelle, E., Moreau, S.*, et al.* (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J 77*, 817-837.

Roy, S., Robson, F., Lilley, J., Liu, C.W., Cheng, X., Wen, J., Walker, S., Sun, J., Cousins, D., Bone, C.*, et al.* (2017). *MtLAX2*, a Functional Homologue of the Arabidopsis Auxin Influx Transporter AUX1, Is Required for Nodule Organogenesis. *Plant Physiol 174*, 326-338.

Ruffel, S., Freixes, S., Balzergue, S., Tillard, P., Jeudy, C., Martin-Magniette, M.L., van der Merwe, M.J., Kakar, K., Gouzy, J., Fernie, A.R.*, et al.* (2008). Systemic signaling of the plant nitrogen status triggers specific transcriptome responses depending on the nitrogen source in *Medicago truncatula*. *Plant Physiol 146*, 2020-2035.

Ruffel, S., Krouk, G., Ristova, D., Shasha, D., Birnbaum, K.D., and Coruzzi, G.M. (2011). Nitrogen economics of root foraging: transitive closure of the nitrate-cytokinin relay and distinct systemic signaling for N supply vs. demand. *Proc Natl Acad Sci* U S A *108*, 18524-18529.

Salome, P.A., and McClung, C.R. (2005). *PSEUDO-RESPONSE REGULATOR 7* and *9* are partially redundant genes essential for the temperature responsiveness of the *Arabidopsis* circadian clock. *Plant Cell 17*, 791-803.

Sato, N., Ohshima, K., Watanabe, A., Ohta, N., Nishiyama, Y., Joyard, J., and Douce, R. (1998). Molecular characterization of the PEND protein, a novel bZIP protein present in the envelope membrane that is the site of nucleoid replication in developing plastids. *Plant Cell 10*, 859-872.

Schafer, P., Pfiffi, S., Voll, L.M., Zajic, D., Chandler, P.M., Waller, F., Scholz, U., Pons-Kuhnemann, J., Sonnewald, S., Sonnewald, U*., et al.* (2009). Manipulation of plant innate immunity and gibberellin as factor of compatibility in the mutualistic association of barley roots with *Piriformospora indica*. Plant J *59*, 461-474.

Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carre, I.A., and Coupland, G. (1998). The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. *Cell 93*, 1219-1229.

Schauser, L., Wieloch, W., and Stougaard, J. (2005). Evolution of NIN-like proteins in *Arabidopsis*, rice, and *Lotus japonicus*. *J Mol Evol 60*, 229-237.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc 3*, 1101-1108.

Schnabel, E., Journet, E.P., de Carvalho-Niebel, F., Duc, G., and Frugoli, J. (2005). The *Medicago truncatula SUNN* gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length. *Plant Mol Biol 58*, 809-822.

Schnabel, E.L., Kassaw, T.K., Smith, L.S., Marsh, J.F., Oldroyd, G.E., Long, S.R., and Frugoli, J.A. (2011). The *ROOT DETERMINED NODULATION1* gene regulates nodule number in roots of *Medicago truncatula* and defines a highly conserved, uncharacterized plant gene family. *Plant Physiol 157*, 328-340.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods 9*, 671-675.

Seabra, A.R., Silva, L.S., and Carvalho, H.G. (2013). Novel aspects of glutamine synthetase (GS) regulation revealed by a detailed expression analysis of the entire GS gene family of *Medicago truncatula* under different physiological conditions. *BMC Plant Biol 13*, 137.

Searle, I.R., Men, A.E., Laniya, T.S., Buzas, D.M., Iturbe-Ormaetxe, I., Carroll, B.J., and Gresshoff, P.M. (2003). Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. *Science 299*, 109-112.

Shikata, M., Matsuda, Y., Ando, K., Nishii, A., Takemura, M., Yokota, A., and Kohchi, T. (2004). Characterization of *Arabidopsis* ZIM, a member of a novel plant-specific GATA factor gene family. *J Exp Bot 55*, 631-639.

Shim, J.S., Kubota, A., and Imaizumi, T. (2017). Circadian Clock and Photoperiodic Flowering in Arabidopsis: CONSTANS Is a Hub for Signal Integration. Plant Physiol *173*, 5-15.

Shimomura, A., Naka, A., Miyazaki, N., Moriuchi, S., Arima, S., Sato, S., Hirakawa, H., Hayashi, M., Maymon, M., Hirsch, A.M.*, et al.* (2016). Blue Light Perception by Both Roots and Rhizobia Inhibits Nodule Formation in *Lotus japonicus*. *Mol Plant Microbe Interact 29*, 786-796.

Shin, J., Heidrich, K., Sanchez-Villarreal, A., Parker, J.E., and Davis, S.J. (2012). TIME FOR COFFEE represses accumulation of the MYC2 transcription factor to provide time-of-day regulation of jasmonate signaling in *Arabidopsis*. *Plant Cell 24*, 2470-2482.

Somers, D.A., Samac, D.A., and Olhoft, P.M. (2003). Recent advances in legume transformation. *Plant Physiol 131*, 892-899.

Song, Y.H., Ito, S., and Imaizumi, T. (2010). Similarities in the circadian clock and photoperiodism in plants. *Curr Opin Plant Biol 13*, 594-603.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell 9*, 3273-3297.

Sprent, J.I., and James, E.K. (2007). Legume evolution: where do nodules and mycorrhizas fit in? *Plant Physiol 144*, 575-581.

Stagnari, F., Maggio, A., Galieni, A., and Pisante, M. (2017). Multiple benefits of legumes for agriculture sustainability: an overview. Chemical and Biological Technologies in Agriculture *4*, 2.

Staley, C., Ferrieri, A.P., Tfaily, M.M., Cui, Y., Chu, R.K., Wang, P., Shaw, J.B., Ansong, C.K., Brewer, H., Norbeck, A.D.*, et al.* (2017). Diurnal cycling of rhizosphere bacterial communities is associated with shifts in carbon metabolism. *Microbiome 5*, 65.

Stepanova, A.N., Yun, J., Likhacheva, A.V., and Alonso, J.M. (2007). Multilevel interactions between ethylene and auxin in *Arabidopsis* roots. *Plant Cell 19*, 2169-2185.

Strabala, T.J., O'Donnell P, J., Smit, A.M., Ampomah-Dwamena, C., Martin, E.J., Netzler, N., Nieuwenhuizen, N.J., Quinn, B.D., Foote, H.C., and Hudson, K.R. (2006). Gain-of-function phenotypes of many *CLAVATA3/ESR* genes, including four new family members, correlate with tandem variations in the conserved CLAVATA3/ESR domain. *Plant Physiol 140*, 1331-1344.

Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Mas, P., Panda, S., Kreps, J.A., and Kay, S.A. (2000). Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. *Science 289*, 768-771.

Streeter, J.G. (1989). Estimation of ammonium concentration in the cytosol of soybean nodules. *Plant Physiol 90*, 779-782.

Sun, L., Ge, Y., Bancroft, A.C., Cheng, X., and Wen, J. (2018). FNBtools: A Software to Identify Homozygous Lesions in Deletion Mutant Populations. *Front Plant Sci 9*, 976.

Suzaki, T., Yano, K., Ito, M., Umehara, Y., Suganuma, N., and Kawaguchi, M. (2012). Positive and negative regulation of cortical cell division during root nodule development in *Lotus japonicus* is accompanied by auxin response. *Development 139*, 3997-4006.

Suzuki, A., Hara, H., Kinoue, T., Abe, M., Uchiumi, T., Kucho, K., Higashi, S., Hirsch, A.M., and Arima, S. (2008). Split-root study of autoregulation of nodulation in the model legume *Lotus japonicus*. *J Plant Res 121*, 245-249.

Swarup, R., Perry, P., Hagenbeek, D., Van Der Straeten, D., Beemster, G.T., Sandberg, G., Bhalerao, R., Ljung, K., and Bennett, M.J. (2007). Ethylene upregulates auxin biosynthesis in *Arabidopsis* seedlings to enhance inhibition of root cell elongation. *Plant Cell 19*, 2186-2196.

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., Cayrel, A., Endre, G., Zhao, P.X., Chabaud, M.*, et al.* (2008). Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. Plant J *54*, 335-347.

Teakle, G.R., Manfield, I.W., Graham, J.F., and Gilmartin, P.M. (2002). *Arabidopsis thaliana* GATA factors: organisation, expression and DNA-binding characteristics. *Plant Mol Biol 50*, 43-57.

Tesfaye, M., Silverstein, K.A., Nallu, S., Wang, L., Botanga, C.J., Gomez, S.K., Costa, L.M., Harrison, M.J., Samac, D.A., Glazebrook, J.*, et al.* (2013). Spatio-temporal expression patterns of Arabidopsis thaliana and *Medicago truncatula* defensin-like genes. *PLoS One 8*, e58992.

Tirichine, L., Sandal, N., Madsen, L.H., Radutoiu, S., Albrektsen, A.S., Sato, S., Asamizu, E., Tabata, S., and Stougaard, J. (2007). A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science 315*, 104-107.

Tiricz, H., Szucs, A., Farkas, A., Pap, B., Lima, R.M., Maroti, G., Kondorosi, E., and Kereszt, A. (2013). Antimicrobial nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in the transcriptome of *Sinorhizobium meliloti*. *Appl Environ Microbiol 79*, 6737-6746.

Tsay, Y.F., Chiu, C.C., Tsai, C.B., Ho, C.H., and Hsu, P.K. (2007). Nitrate transporters and peptide transporters. *FEBS Lett 581*, 2290-2300.

Tyerman, S.D., Whitehead, L.F., and Day, D.A. (1995). A channel-like transporter for NH4+ on the symbiotic interface of N2-fixing plants. *Nature 378*, 629.

Udvardi, M., and Poole, P.S. (2013). Transport and metabolism in legume-rhizobia symbioses. *Annu Rev Plant Biol 64*, 781-805.

Udvardi, M.K., and Day, D.A. (1997). Metabolite Transport across Symbiotic Membranes of Legume Nodules. *Annu Rev Plant Physiol Plant Mol Biol 48*, 493-523.

Ueoka-Nakanishi, H., Yamashino, T., Ishida, K., Kamioka, M., Nakamichi, N., and Mizuno, T. (2012). Molecular Mechanisms of Circadian Rhythm in *Lotus japonicus* and *Arabidopsis thaliana* Are Sufficiently Compatible to Regulate Heterologous

Core Clock Genes Robustly. *Bioscience, Biotechnology, and Biochemistry 76*, 2332-2334.

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., and Vandepoele, K. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res 46*, D1190-D1196.

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., Tiricz, H.*, et al.* (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science 327*, 1122-1126.

van Noorden, G.E., Ross, J.J., Reid, J.B., Rolfe, B.G., and Mathesius, U. (2006). Defective long-distance auxin transport regulation in the *Medicago truncatula* super numeric nodules mutant. *Plant Physiol 140*, 1494-1506.

Vernie, T., Kim, J., Frances, L., Ding, Y., Sun, J., Guan, D., Niebel, A., Gifford, M.L., de Carvalho-Niebel, F., and Oldroyd, G.E. (2015). The NIN Transcription Factor Coordinates Diverse Nodulation Programs in Different Tissues of the *Medicago truncatula* Root. *Plant Cell 27*, 3410-3424.

Vernie, T., Moreau, S., de Billy, F., Plet, J., Combier, J.P., Rogers, C., Oldroyd, G., Frugier, F., Niebel, A., and Gamas, P. (2008). EFD Is an ERF transcription factor involved in the control of nodule number and differentiation in *Medicago truncatula*. *Plant Cell 20*, 2696-2713.

Vessey, J.K., and Waterer, J. (1992). In search of the mechanism of nitrate inhibition of nitrogenase activity in legume nodules: Recent developments. Physiologia Plantarum *84*, 171-176.

Vidal, E.A., Alvarez, J.M., Moyano, T.C., and Gutierrez, R.A. (2015). Transcriptional networks in the nitrate response of *Arabidopsis thaliana*. Curr Opin Plant Biol *27*, 125-132.

Vidal, E.A., and Gutierrez, R.A. (2008). A systems view of nitrogen nutrient and metabolite responses in *Arabidopsis*. *Curr Opin Plant Biol 11*, 521-529.

Vinardell, J.M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Kelemen, Z., Tarayre, S., Roudier, F., Mergaert, P., Kondorosi, A.*, et al.* (2003). Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. *Plant Cell 15*, 2093-2105.

Vincent, J.M. (1970). A manual for the practical study of root-nodule bacteria (Oxford: Published for the International Biological Programme by Blackwell Scientific).

Voss, U., Wilson, M.H., Kenobi, K., Gould, P.D., Robertson, F.C., Peer, W.A., Lucas, M., Swarup, K., Casimiro, I., Holman, T.J.*, et al.* (2015). The circadian clock rephases during lateral root organ initiation in *Arabidopsis thaliana*. *Nat Commun 6*, 7641.

Wang, D., Griffitts, J., Starker, C., Fedorova, E., Limpens, E., Ivanov, S., Bisseling, T., and Long, S. (2010). A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science 327*, 1126-1129.

Wang, Q., Yang, S., Liu, J., Terecskei, K., Abraham, E., Gombar, A., Domonkos, A., Szucs, A., Kormoczi, P., Wang, T.*, et al.* (2017). Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. *Proc Natl Acad Sci* U S A *114*, 6854-6859.

Wang, R., Okamoto, M., Xing, X., and Crawford, N.M. (2003a). Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol 132*, 556-567.

Wang, R., Tischner, R., Gutierrez, R.A., Hoffman, M., Xing, X., Chen, M., Coruzzi, G., and Crawford, N.M. (2004). Genomic analysis of the nitrate response using a nitrate reductase-null mutant of *Arabidopsis*. *Plant Physiol 136*, 2512-2522.

Wang, R., Xing, X., and Crawford, N. (2007). Nitrite acts as a transcriptome signal at micromolar concentrations in *Arabidopsis* roots. *Plant Physiol 145*, 1735-1745.

Wang, W., Barnaby, J.Y., Tada, Y., Li, H., Tor, M., Caldelari, D., Lee, D.U., Fu, X.D., and Dong, X. (2011). Timing of plant immune responses by a central circadian regulator. *Nature 470*, 110-114.

Wang, Y., Henriksson, E., Soderman, E., Henriksson, K.N., Sundberg, E., and Engstrom, P. (2003b). The *Arabidopsis* homeobox gene, *ATHB16*, regulates leaf development and the sensitivity to photoperiod in *Arabidopsis*. *Dev Biol 264*, 228-239.

Wang, Y.Y., Hsu, P.K., and Tsay, Y.F. (2012). Uptake, allocation and signaling of nitrate. *Trends Plant Sci 17*, 458-467.

Wang, Z.Y., and Tobin, E.M. (1998). Constitutive expression of the *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)* gene disrupts circadian rhythms and suppresses its own expression. Cell *93*, 1207-1217.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics 25*, 1189-1191.

Weigel, D., and Glazebrook, J. (2008). Genetic analysis of *Arabidopsis* mutants. *CSH Protoc 2008*, pdb top35.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K.*, et al.* (2014). Determination

and inference of eukaryotic transcription factor sequence specificity. *Cell 158*, 1431-1443.

Wopereis, J., Pajuelo, E., Dazzo, F.B., Jiang, Q., Gresshoff, P.M., De Bruijn, F.J., Stougaard, J., and Szczyglowski, K. (2000). Short root mutant of *Lotus japonicus* with a dramatically altered symbiotic phenotype. *Plant J 23*, 97-114.

Wu, J.L., Wu, C., Lei, C., Baraoidan, M., Bordeos, A., Madamba, M.R., Ramos-Pamplona, M., Mauleon, R., Portugal, A., Ulat, V.J.*, et al.* (2005). Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Mol Biol 59*, 85-97.

Xu, C., Luo, F., and Hochholdinger, F. (2016). LOB Domain Proteins: Beyond Lateral Organ Boundaries. *Trends Plant Sci 21*, 159-167.

Yakir, E., Hilman, D., Harir, Y., and Green, R.M. (2007). Regulation of output from the plant circadian clock. *FEBS J 274*, 335-345.

Yamashino, T., Ito, S., Niwa, Y., Kunihiro, A., Nakamichi, N., and Mizuno, T. (2008). Involvement of *Arabidopsis* clock-associated pseudo-response regulators in diurnal oscillations of gene expression in the presence of environmental time cues. *Plant Cell Physiol 49*, 1839-1850.

Yang, S., Wang, Q., Fedorova, E., Liu, J., Qin, Q., Zheng, Q., Price, P.A., Pan, H., Wang, D., Griffitts, J.S.*, et al.* (2017). Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. *Proc Natl Acad Sci* U S A *114*, 6848-6853.

Yazdanbakhsh, N., Sulpice, R., Graf, A., Stitt, M., and Fisahn, J. (2011). Circadian control of root elongation and C partitioning in *Arabidopsis thaliana*. *Plant Cell Environ 34*, 877-894.

Yendrek, C.R., Lee, Y.C., Morris, V., Liang, Y., Pislariu, C.I., Burkart, G., Meckfessel, M.H., Salehin, M., Kessler, H., Wessler, H.*, et al.* (2010). A putative transporter is essential for integrating nutrient and hormone signaling with lateral root growth and nodule development in *Medicago truncatula*. *Plant J 62*, 100-112.

Yoshida, C., Funayama-Noguchi, S., and Kawaguchi, M. (2010). plenty, a novel hypernodulation mutant in *Lotus japonicus*. *Plant Cell Physiol 51*, 1425-1435.

Young, M.W., and Kay, S.A. (2001). Time zones: a comparative genetics of circadian clocks. *Nat Rev Genet 2*, 702-715.

Young, N.D., Debelle, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., Schoof, H.*, et al.* (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature 480*, 520-524.

Yu, N., Niu, Q.W., Ng, K.H., and Chua, N.H. (2015). The role of miR156/SPLs modules in *Arabidopsis* lateral root development. *Plant J 83*, 673-685.

Zhang, C., Hou, Y., Hao, Q., Chen, H., Chen, L., Yuan, S., Shan, Z., Zhang, X., Yang, Z., Qiu, D.*, et al.* (2015). Genome-wide survey of the soybean GATA transcription factor gene family and expression analysis under low nitrogen stress. *PLoS One 10*, e0125174.

Zhang, C., Xie, Q., Anderson, R.G., Ng, G., Seitz, N.C., Peterson, T., McClung, C.R., McDowell, J.M., Kong, D., Kwak, J.M.*, et al.* (2013). Crosstalk between the circadian clock and innate immunity in *Arabidopsis*. *PLoS Pathog 9*, e1003370.

Zhang, H., Jennings, A., Barlow, P.W., and Forde, B.G. (1999). Dual pathways for regulation of root branching by nitrate. *Proc Natl Acad Sci* U S A *96*, 6529-6534.

Zhao, C.Z., Huang, J., Gyaneshwar, P., and Zhao, D. (2017). *Rhizobium* sp. IRBG74 Alters *Arabidopsis* Root Development by Affecting Auxin Signaling. *Front Microbiol 8*, 2556.

Zhao, Y., Medrano, L., Ohashi, K., Fletcher, J.C., Yu, H., Sakai, H., and Meyerowitz, E.M. (2004). *HANABA TARANU* is a GATA transcription factor that regulates shoot apical meristem and flower development in *Arabidopsis*. *Plant Cell 16*, 2586-2600.

Zheng, B., Deng, Y., Mu, J., Ji, Z., Xiang, T., Niu, Q.-W., Chua, N.-H., and Zuo, J. (2006). Cytokinin affects circadian-clock oscillation in a phytochrome B- and *Arabidopsis* response regulator 4-dependent manner. *Physiologia Plantarum 127*, 277-292.